Approximate Diagnosis and Opacity of Stochastic Systems

Engel Lefaucheux
Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany
elefauch@mpi-sws.org

Abstract

We consider the control of the information released by a system represented by a stochastic model. In this framework, an external observer is interested in detecting a particular set of relevant paths of the system. However, he can only observe those paths trough an observation function which obfuscates the real behaviour of the system. Exact disclosure occurs when the observer can deduce from a finite observation that the path is relevant, the approximate disclosure variant corresponding to the path being identified as relevant with arbitrarily high accuracy. We consider the problems of diagnosability and opacity, which corresponds, in spirit, to the cases where one wants to disclose all the information or hide as much of it as possible. While these problems have already been studied for the exact disclosure notion, there are very few works considering the approximate disclosure. Under the approximate notion of disclosure, we establish that opacity of Markov chains is in EXPTIME and PSPACE-hard. Moreover, we show that diagnosability is EXPTIME-complete for controllable systems while nearly every opacity question is undecidable in an active setting.

2012 ACM Subject Classification General and reference → Verification; Theory of computation → Probabilistic computation

Keywords and phrases Stochastic systems, Opacity, Diagnosis

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

Diagnosis and Opacity

Due to the omnipresence of communicating devices, controlling the information produced by a system has seen an increasing importance. This control mainly takes two directions. First, it can be done in order to detect some internal behaviour, such as malfunctions, of the system. In control theory, this direction has been formalised under the name diagnosis and studied on systems modelled by partially observable labelled transition systems (POLTS) [25]. In such a framework, diagnosability requires that the occurrence of unobservable faults can be deduced accurately from the previous and subsequent observable events. Diagnosability for POLTS was shown to be decidable in PTIME [19]. Also, several contributions, gathered under the generic name of active diagnosis, focus on enforcing the diagnosability of a system [24, 27, 14, 15]. The second direction of information control aims at hiding a secret behaviour of the system. This property, called opacity, is motivated by security: an external user should not, by observing an execution of a system, acquire the guarantee that it is a secret one. This property was formalised for POLTS [13] by specifying a subset of secret paths and requiring that, for any secret path, there is a non-secret one with the same observation. Both diagnosability and opacity thus consider a set of relevant paths. The *disclosure set* of a system is then the set of relevant paths that can be identified as such.

Information control of stochastic systems

In stochastic systems, one can use the probabilities to refine the analysis of the disclosure set. First, probabilities allow to quantify the importance of the leak of information. In

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

this endeavour, various measures for the disclosure set, called probabilistic disclosure, were introduced [23, 1, 5, 3]. Secondly, in stochastic systems, the ability to identify a path as relevant can be chosen to depend on the probabilities. There are three natural variants: (1) exact disclosure, which, as in the non-stochastic case, require that no non-relevant path share the same observation, (2) ε-disclosure for $\varepsilon > 0$ which tolerates small errors, allowing to claim the relevance of a path if the conditional probability that the path is relevant exceeds $1 - \varepsilon$, and (3) Accurate Approximate disclosure (AA-disclosure) which is satisfied when the accuracy of the guess can be chosen arbitrarily high. Under the exact notion of disclosure, both diagnosability and opacity have been studied extensively for stochastic systems [7, 6, 26, 2, 4]. In particular, various exact notions of diagnosability have been shown to be PSPACE-complete for observable Markov chains (oMC). The study of the approximate notions of disclosure has however been more limited, especially for the notion of AA-disclosure. The most notable result showed that diagnosability under AA-disclosure is decidable in PTIME [8] for oMC.

Contribution

In this paper, we study diagnosability and opacity in stochastic systems under AA-disclosure.

- we formally introduce a notion of accurate approximate opacity that mirrors the existing diagnosability notion (Definition 7);
- we show that opacity with AA-disclosure for oMC is PSPACE-complete (Theorem 11);
- we establish that diagnosability with AA-disclosure for weighted Markov chains, a controllable setting, is EXPTIME-complete (Theorem 17);
- we prove the undecidability of most notions of opacity under AA-disclosure for observable Markov decision processes (Theorem 26 and 28).

Organisation

In Section 2, we define and discuss the notions of disclosure, diagnosability as well as opacity. We also gather and complete the results on approximate diagnosability and opacity in a passive framework. Then, in Section 3, we consider diagnosability for weighted Markov chains, a framework giving a partial external control on the system. Similarly, in Section 4, we study opacity for observable Markov decision processes, a setting where the control is more powerful than the one in weighted Markov chains due to being internal to the system. For space concerns, the most technical proofs are deferred to the appendix.

2 Diagnosis and Opacity for Markov Chains

2.1 Observable Markov Chains

For a finite alphabet Σ, we denote by Σ^* (resp. Σ^ω) the set of finite (resp. infinite) words over Σ, $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$ and ε the empty word. The length of a word w is denoted by $|w| \in \mathbb{N} \cup \{\infty\}$ and for $n \in \mathbb{N}$, Σ^n is the set of words of length n. A word $u \in \Sigma^*$ is a prefix of $v \in \Sigma^\infty$, written $u \leq v$, if $v = uw$ for some $w \in \Sigma^\infty$. The prefix is strict if $w \neq \varepsilon$. For $n \leq |w|$, we write $w_{\downarrow n}$ for the prefix of length n of w. Given a countable set S, a distribution on S is a mapping $\mu : S \to [0, 1]$ such that $\sum_{s \in S} \mu(s) = 1$. The support of μ is $\mathsf{Supp}(\mu) = \{s \in S \mid \mu(s) > 0\}$. If $\mathsf{Supp}(\mu) = \{s\}$ is a single element, μ is a Dirac distribution on s written $\mathbf{1}_s$. We denote by $\mathsf{Dist(S)}$ the set of distributions on S.

For the purpose of information control questions, the model must be equipped with an *observation function* describing what an external observer can see. The observation function can be obtained via a labelling of states or transitions, both options being known to be equivalent. We thus define observable Markov chains (see Figure 1).

▶ **Definition 1** (Observable Markov chains)**.** *An observable Markov chain (oMC) over alphabet Σ is a tuple $\mathcal{M} = (S, p, \mathsf{O})$ where S is a countable set of states, $p : S \to \mathsf{Dist(S)}$ is the transition function, and $\mathsf{O} : S \to \Sigma$ is the observation function.*

We write $p(s'|s)$ instead of $p(s)(s')$ to emphasise the probability of going to state s' conditioned by being in state s. Given a distribution $\mu_0 \in \mathsf{Dist(S)}$, we denote by $\mathcal{M}(\mu_0)$ the oMC with initial distribution μ_0. For decidability and complexity results, we assume that all probabilities occurring in the model (transition probabilities and initial distribution) are rationals. A (finite or infinite) path of $\mathcal{M}(\mu_0)$ is a sequence of states $\rho = s_0 s_1 \ldots \in S^\infty$ such that $\mu_0(s_0) > 0$ and for each $i \geq 0$, $p(s_{i+1}|s_i) > 0$. For a finite path, $\rho = s_0 s_1 \ldots s_n$, we call n its length and denote its ending state by $\mathsf{last}(\rho) = s_n$. A finite path ρ_1 prefixes a finite or infinite path ρ if there exists a path ρ_2 such that $\rho = \rho_1 \rho_2$. The set $\mathsf{Cyl}(\rho)$ represents the cylinder of infinite paths prefixed by ρ. We denote by $\mathsf{Path}(\mathcal{M}(\mu_0))$ (resp. $\mathsf{FPath}(\mathcal{M}(\mu_0))$) the set of infinite (finite) paths of $\mathcal{M}(\mu_0)$. The *observation sequence* of the path $\rho = s_0 s_1 \ldots$ is the word $\mathsf{O}(\rho) = \mathsf{O}(s_0)\mathsf{O}(s_1)... \in \Sigma^\infty$. For a set R of paths, $\mathsf{O}(R) = \{\mathsf{O}(\rho) \mid \rho \in R\}$ and for a set W of observation sequences, $\mathsf{O}^{-1}(W) = \{\rho \mid \mathsf{O}(\rho) \in W\}$.

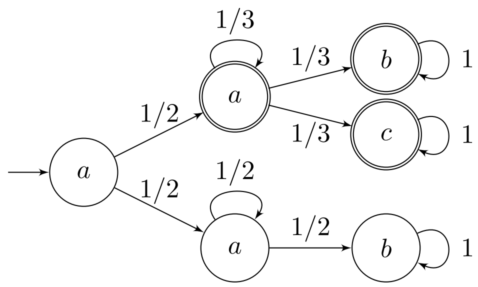

Figure 1 An observable Markov chain with disclosure $\frac{1}{4}$. The arrow entering the leftmost state means that the initial disribution is a Dirac on this state. Relevant states are circled twice.

Forgetting the labels, an oMC with an initial distribution μ_0 becomes a discrete time Markov chain (DTMC). In a DTMC, the set of infinite paths is the support of a probability measure extended from the probabilities of the cylinders by the Caratheodory's extension theorem:

$$\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{Cyl}(s_0 s_1 \ldots s_n)) = \mu_0(s_0) p(s_1|s_0) \ldots p(s_n|s_{n-1}) \ .$$

When $\mathcal{M}(\mu_0)$ is clear from context, we will sometimes omit the subscript, and write $\mathbf{P}$ for $\mathbf{P}_{\mathcal{M}(\mu_0)}$. Let $\rho \in \mathsf{FPath}(\mathcal{M})$, $w \in \Sigma^*$ and $E \subseteq \Sigma^\omega$, with a small abuse of notation we write $\mathbf{P}(\rho)$ for $\mathbf{P}(\mathsf{Cyl}(\rho))$, $\mathbf{P}(w)$ instead of $\mathbf{P}(\cup_{\rho \in \mathsf{O}^{-1}(w)} \mathsf{Cyl}(\rho))$ and $\mathbf{P}(E)$ instead of $\mathbf{P}(\{\rho \in \mathsf{Path}(\mathcal{M}(\mu_0)) \mid \rho \in \mathsf{O}^{-1}(E)\})$.

2.2 Relevant Paths and Notions of Disclosure

In this paper, we consider diagnosability and opacity problems; in both cases, one needs to identify a set of paths of the system which carries hidden information (the paths represent a faulty behavior of the system for diagnosability, and they represent a secret one for opacity). We focus on the particular case where the relevant behavior of the system is given by a subset of states $\mathsf{S}^\mathsf{r} \subseteq S$, called *relevant states*, of the model: a (finite or infinite) path $s_0 s_1 \dots$ is *relevant* if $s_i \in \mathsf{S}^\mathsf{r}$ for some i. The set of infinite relevant paths is denoted Rel.

▶ Remark 2. Without loss of generality, we can assume that the set of relevant states is absorbing (see [2] for example).

In stochastic systems, the set of paths disclosing that they are relevant depends on the level of confidence that the observer wants. To measure this, we define the proportion of relevant paths among those having the same observation sequence as follow:

▶ **Definition 3** (Proportion of relevant paths)**.** *Given an oMC $\mathcal{M} = (S, p, \mathsf{O})$, an initial distribution μ_0, $\mathsf{S}^\mathsf{r} \subseteq S$ and an observation sequence $w \in \Sigma^*$, the proportion of relevant paths associated with the observation sequence w is:*

$$\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w) = \frac{\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(w) \mid \rho \in \mathsf{Rel}\})}{\mathbf{P}(w)}.$$

▶ **Example 4.** Consider the oMC of Figure 1 and the observation sequences a^k, $a^k b^n$ and $a^k c^m$. The observation sequence a^k, for $k > 1$, can be produced by a non-relevant path with probability $1/2^{k-1}$ and by a relevant path with probability $1/2 \times 1/3^{k-2}$. Therefore, $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(a^k) = \frac{1/3^{k-2}}{1/2^{k-2}+1/3^{k-2}}$ which converges to 0 when k grows to infinity. The proportion of relevant paths for the observation $a^k b^n$ with $k > 1$ and $n \geq 1$ is similarly $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(a^k b^n) = \frac{1/2^{k-1}}{1/2^{k-1}+1/3^{k-1}}$ which remains constant for extensions of $a^k b^n$ as it does not depend on n. Finally, if $m \geq 1$, $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(a^k c^m) = 1$ as no non-relevant path can produce a 'c'.

Using this proportion, we define a measure on the quantity of information disclosed by a system. We first introduce a notion of approximate *disclosure* where one considers that a path reveals its relevance if the proportion of relevant paths of its observation sequence is greater than $1 - \varepsilon$ for some given $\varepsilon > 0$.

▶ **Definition 5** (Approximate information control)**.** *Given an oMC $\mathcal{M} = (S, p, \mathsf{O})$, an initial distribution μ_0, $\mathsf{S}^\mathsf{r} \subseteq S$ and $\varepsilon > 0$, an observation sequence $w \in \Sigma^*$ is ε-disclosing if $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w) > 1 - \varepsilon$. It is ε-min-disclosing if it is ε-disclosing and no strict prefix of w is ε-disclosing. Writing $D^\varepsilon_{\min}$ for the set of ε-min-disclosing observation sequences, the ε-disclosure is defined by*

$$Disc^\varepsilon(\mathcal{M}(\mu_0)) = \sum_{w \in D^\varepsilon_{\min}} \mathbf{P}(\{\rho \in \mathsf{Rel} \mid \exists \rho' \leq \rho, \mathsf{O}(\rho') = w\})$$

This definition raises the two following decision problems for any $0 \leq \varepsilon < 1$:

- **For opacity:** the ε-disclosure problem consists in, given $\lambda \in [0;1]$, deciding whether $Disc^\varepsilon(\mathcal{M}(\mu_0)) > \lambda$.
- **For diagnosis:** the ε-diagnosability problem consists in deciding whether $Disc^\varepsilon(\mathcal{M}(\mu_0)) = \mathbf{P}(\mathsf{Rel})$.

We can see an asymmetry between the problems introduced for opacity and for diagnosis here: in the former the threshold the ε-disclosure is compared to is given by the user while in the latter it is derived from the system. The reason for this difference is that a failure of the system is often considered important and must be detected, while a small chance of leaking information may be deemed acceptable. Unfortunately, it is known that these problems are undecidable for $\varepsilon \neq 0$ [1]:

[1] The case $\varepsilon = 0$, with a non-strict inequality, is a form of exact disclosure for which some problems are decidable [2, 9].

▶ **Theorem 6.** *Given $0 < \varepsilon < 1$, the positive ε-disclosure problem [2] and the ε-diagnosability problem [8] are undecidable for oMCs.*

In order to regain decidability one can consider slightly more qualitative notions of approximate information control, that we call accurate approximate. Instead of deeming the relevance of a path to be revealed when the proportion of relevant path goes above a given threshold, an infinite observation sequence is AA-*disclosing* if this proportion converges toward 1. In other words, when observing an AA-disclosing observation sequence, by waiting, one can get an arbitrarily high confidence that the path is relevant.

▶ **Definition 7** (Accurate approximate information control). *Given an oMC $\mathcal{M} = (S, p, \mathsf{O})$, an initial distribution μ_0, and $\mathsf{S^r} \subseteq S$, an observation sequence $w \in \Sigma^\omega$ is AA-disclosing if $\lim_{n\to\infty} \mathsf{P^{rel}}_{\mathcal{M}(\mu_0)}(w_{\downarrow n}) = 1$. Writing D^{AA} for the set of AA-disclosing observation sequences, the AA-disclosure is defined by*

$$Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = \sum_{w \in D^{\mathsf{AA}}} \mathbf{P}(\{\rho \in Rel \mid \mathsf{O}(\rho) = w\})$$

As before, this definition raises two decision problems:

- **For opacity:** the AA-disclosure problem consists in, given $\lambda \in [0;1]$, deciding if $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) > \lambda$.
- **For diagnosis:** the AA-diagnosability problem consists in deciding if $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = \mathbf{P}(\mathsf{Rel})$.

AA-diagnosability was initially defined in [26] slightly differently: a system was called AA-diagnosable if it was ε-diagnosable for all $\varepsilon > 0$. We introduced the new definition with the study of active systems in mind. However, the two definitions are in fact equivalent for oMC.

▶ **Proposition 8.** *An oMC is AA-diagnosable iff it is ε-diagnosable for all $\varepsilon > 0$.*

2.3 Decidability of the Accurate Approximate Problems for oMCs

With the accurate approximate approach to information control, one regain decidability. The AA-diagnosability problem for finite oMC was shown to be in PTIME in [8]. This result relies on the notion of distance between two oMC introduced in [17] and defined in the following way: the distance between two oMC $\mathcal{M}_1$ and $\mathcal{M}_2$ with initial distribution μ_1 and μ_2 is

$$d(\mathcal{M}_1(\mu_1), \mathcal{M}_2(\mu_2)) = \max_{E \subseteq \Sigma^\omega} \mathbf{P}_{\mathcal{M}_1(\mu_1)}(E) - \mathbf{P}_{\mathcal{M}_2(\mu_2)}(E).$$

The authors of [17] show how to decide in PTIME if the distance between two oMC is 1 thanks to the following characterisation.

▶ **Proposition 9** ([17]). *Given two oMC $\mathcal{M}_1$ and $\mathcal{M}_2$ and two initial distributions μ_1 and μ_2, $d(\mathcal{M}_1(\mu_1), \mathcal{M}_2(\mu_2)) < 1$ iff there exists $w \in \Sigma^*$ and two distributions π_1 and π_2 such that, writing μ_1^w and μ_2^w for the probability distributions reached after observing w in $\mathcal{M}_1(\mu_1)$ and $\mathcal{M}_2(\mu_2)$ respectively, we have, for $i \in \{1,2\}$, $\mathsf{Supp}(\pi_i) \subseteq \mathsf{Supp}(\mu_i^w)$ and $d(\mathcal{M}_1(\pi_1), \mathcal{M}_2(\pi_2)) = 0$ (*i.e. $\forall w' \in \Sigma^*, \mathbf{P}_{\mathcal{M}_1(\pi_1)}(w') = \mathbf{P}_{\mathcal{M}_2(\pi_2)}(w')$*).*

Finally, the link between the distance 1 of two oMC and AA-diagnosability was established in [8], giving the PTIME algorithm:

▶ **Theorem 10** ([8]). *Let $\mathcal{M}$ be a finite oMC and μ_0 be an initial distribution. $\mathcal{M}(\mu_0)$ is not* AA*-diagnosable iff there exist two states $s \in \mathsf{S}^\mathsf{r}$ and $s' \in S \setminus \mathsf{S}^\mathsf{r}$ with s' belonging to a bottom strongly connected component (BSCC) of $\mathcal{M}$ and there exist two finite paths ρ and ρ' of* $\mathsf{FPath}(\mathcal{M}(\mu_0))$ *such that* $\mathsf{last}(\rho) = s$, $\mathsf{last}(\rho') = s'$, $\mathsf{O}(\rho) = \mathsf{O}(\rho')$ *and* $d(\mathcal{M}(\mathbf{1}_s), \mathcal{M}(\mathbf{1}_{s'})) < 1$.

Considering only the sufficient condition, a more general result allowing for infinite oMC and which we will need later, was in fact proven in [8]: Let $\mathcal{M}$ be a (potentially infinite) oMC, μ_0 be an initial distribution, two states $s \in \mathsf{S}^\mathsf{r}$ and $s' \in S \setminus \mathsf{S}^\mathsf{r}$ with s' such that no relevant state can be reached from s' and two finite paths ρ and ρ' of $\mathsf{FPath}(\mathcal{M}(\mu_0))$ such that $\mathsf{last}(\rho) = s$, $\mathsf{last}(\rho') = s'$, $\mathsf{O}(\rho) = \mathsf{O}(\rho')$. Then $\mathcal{M}(\mu_0)$ is AA-diagnosable implies that $d(\mathcal{M}(\mathbf{1}_q), \mathcal{M}(\mathbf{1}_{q'})) = 1$.

While AA-diagnosability can be decided in polynomial time, the AA-disclosure problem is a bit more complicated.

▶ **Theorem 11.** *The* AA*-disclosure problem for finite oMC is* PSPACE*-complete.*

Sketch of proof. In order to solve the AA-disclosure problem in EXPTIME. We first build an exponential size oMC which contains additional information compared to the original one. Then we show that there are two kinds of BSCC in this new oMC: the ones that are reached by paths that almost surely have an AA-disclosing observation sequence, and the ones that are reached by paths that do not correspond to AA-disclosing observation sequences. We then use the existing results for the AA-diagnosability problem to determine the status of each BSCC. Finally, computing the AA-disclosure of the oMC is equivalent to computing the probability to reach the "AA-disclosing" BSCC, which can be done in NC in the size of the oMC, thus giving an overall PSPACE algorithm.

The hardness is obtained by reduction from the universality problem for non-deterministic finite automaton (NFA), which is known to be PSPACE-complete [20]. ◀

3 Active Approximate Diagnosis

We will now consider an active setting where a controller can modify the behaviour of the system. Exact notions of diagnosis [6] and opacity [2] have been studied in an active setting. The frameworks used for each notion are not equivalent however as they do not give the same power to the controller. This difference comes from an intrinsic distinction between the two problems:

- Diagnosability corresponds to situations where one wants to obtain information from the system through exterior control. Therefore the controller is supposed to have the same amount of information as the diagnoser.
- For opacity on the contrary, the control either aims to diffuse an information outside of the system (case of a virus for example) or is implemented in the system during the design to protect it. In these two cases, the controller knows the exact state of the system.

Therefore we consider Weighted Markov Chains to study diagnosability and in the next section we will use Markov Decision Processes for opacity.

3.1 Diagnosis for Weighted Markov chains

▶ **Definition 12** (WMC). *A weighted Markov Chain (WMC) over alphabet Σ is a tuple $\mathbb{M} = (S, T, \mathsf{O})$ where S is a finite set of states, $T : S \times S \to \mathbb{N}$ is the transition function labelling transitions with integer weights and $\mathsf{O} : S \to \Sigma$ is the observation function.*

The alphabet is partitioned into controllable and uncontrollable events $\Sigma = \Sigma_c \uplus \Sigma_e$. A set $\Sigma_0^s \subseteq \Sigma$ of *allowed events* in a state $s \in S$ is a set of observations such that $\Sigma_e \subseteq \Sigma_0^s$ and $\{s' \in S \mid T(s,s') > 0 \wedge \mathsf{O}(s') \in \Sigma_0^s\} \neq \emptyset$. Given a state s and a set of allowed events Σ_0^s, we define the transition probability $p(s, \Sigma_0^s)$ such that for all s' with $\mathsf{O}(s') \in \Sigma_0^s$, $p(s,\Sigma_0^s)(s') = \frac{T(s,s')}{\sum_{s'', \mathsf{O}(s'') \in \Sigma_0^s} T(s,s'')}$. As before, we write $p(s'|s,\Sigma_0^s)$ instead of $p(s,\Sigma_0^s)(s')$. Given an initial distribution μ_0, an infinite path of a WMC $\mathbb{M}$ is a sequence $\rho = s_0\Sigma_0 s_1 \Sigma_1 \dots$ where $\mu_0(s_0) > 0$ and $p(s_{i+1}|s_i,\Sigma_i) > 0$, for $s_i \in S$ and Σ_i is a set of allowed events in s_i, for all $i \geq 0$. As for oMC, we define finite paths, and we use similar notations for the various sets of paths. A sequence of observations and set of allowed events $b \in (\Sigma \times 2^\Sigma)^*\Sigma$ is called a *knowledge sequence*. The knowledge sequence of a path of a WMC $\rho = s_0\Sigma_0 s_1\Sigma_1 \dots s_i$ is $K(\rho) = \mathsf{O}(s_0)\Sigma_0\mathsf{O}(s_1)\Sigma_1 \dots \mathsf{O}(s_i)$.

The nondeterministic choice of the set of allowed events is resolved by strategies.

▶ **Definition 13** (Strategy for WMC). *A* strategy *of WMC $\mathbb{M}$ with initial distribution μ_0 is a mapping $\sigma : (\Sigma \times 2^\Sigma)^*\Sigma \to \mathsf{Dist}(2^\Sigma)$ associating to any knowledge sequence a distribution on sets of events.*

We will only consider here strategies that do not generate a deadlock, i.e. strategies σ such that for all state s reached after a knowledge b, $\sigma(b)$ is a set of allowed events for s. Given a strategy σ, a path $\rho = s_0\Sigma_0 s_1\Sigma_1 \dots$ of $\mathbb{M}$ is *σ-compatible* if for all i, $\Sigma_i \in \mathsf{Supp}(\sigma(K(s_0\Sigma_0 s_1 \Sigma_1 \dots s_i))$. A strategy σ is *deterministic* if $\sigma(b)$ is a Dirac distribution for each knowledge sequence b. In this case, we denote by $\sigma(b)$ the set of allowed actions $\Sigma_a \in 2^\Sigma$ such that $\sigma(b) = \mathbf{1}_{\Sigma_a}$. Let b be a knowledge sequence. We define $B_{\mathbb{M}(\mu_0)}(b)$ the *belief* about states corresponding to b as follows:

$$B_{\mathbb{M}(\mu_0)}(b) = \{s \mid \exists \rho \in \mathsf{FPath}(\mathbb{M}(\mu_0)),\ K(\rho) = b \wedge s = \mathsf{last}(\rho)\}$$

A strategy σ is *belief-based* if for all b, $\sigma(b)$ only depends on its belief $B_{\mathbb{M}(\mu_0)}(b)$ (*i.e.* given two knowledge sequence b and b' if $B_{\mathbb{M}(\mu_0)}(b) = B_{\mathbb{M}(\mu_0)}(b')$ then $\sigma(b) = \sigma(b')$). For belief-based strategies, we will sometimes write $\sigma(B)$ for the choice of the strategy made for knowledge sequences producing the belief B.

As for oMC, the secret is defined by the reachability of a set $\mathsf{S}^\mathsf{r} \in S$ of secret states of the WMC and note that the construction ensuring that once a secret state is visited, the path remains secret forever, extends naturally from oMC to WMC. We consider only WMC of this form in the following.

A strategy σ on $\mathbb{M}(\mu_0)$ defines an infinite Markov chain $\mathbb{M}_\sigma(\mu_0)$ with set of states the finite σ-compatible paths, that can be equipped with the observation function associating $\Sigma_{n-1}\mathsf{O}(s_n)$ with the state associated to the finite path $\rho = s_0\Sigma_0 \dots \Sigma_{n-1}s_n$ (Σ_{n-1} being omitted if $n = 0$). The transition function p_σ is defined for ρ a σ-compatible path and $\rho' = \rho\Sigma_a s'$ by $p_\sigma(\rho'|\rho) = \sigma(K(\rho))(\Sigma_a)p(s'|s,\Sigma_a)$ and we denote by $\mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}$ the associated probability measure. When the strategy possesses some good regularity properties, this oMC is equivalent to a finite one (*i.e.* there is a one-to-one correspondence between the paths of each oMC, it preserves the knowledge sequence and the probability. The two oMC have therefore the same disclosure properties). For instance given a deterministic belief based strategy σ, one can define the oMC $\mathbb{M}'_\sigma$ with set of states $S \times 2^\Sigma \times 2^S$, observation $\mathsf{O}'_\sigma(s, \Sigma^\bullet, B) = (\mathsf{O}(s), \Sigma^\bullet)$, initial distribution $\mu_0^\sigma(s, \emptyset, \mathsf{Supp}(\mu_0) \cap \mathsf{O}^{-1}(\mathsf{O}(s))) = \mu_0(s)$ and transition function $p'_\sigma((s_1,\Sigma_1,B_1) \mid (s_2,\Sigma_2,B_2)) = p(s_2 \mid s_1, \Sigma_2)$ if $\sigma(B_1) = \Sigma_2$ and $B_2 = B_{\mathbb{M}(\mu_1)}(\mathsf{O}(s_2))$ for μ_1 a distribution of support B_1, $p'_\sigma((s_1,\Sigma_1,B_1) \mid (s_2,\Sigma_2,B_2)) = 0$ otherwise. The oMC $\mathbb{M}'_\sigma$ is exponential in the size of $\mathbb{M}$ and is equivalent to $\mathbb{M}_\sigma$. When considering belief-based strategies, we will call $\mathbb{M}_\sigma$ the finite equivalent oMC.

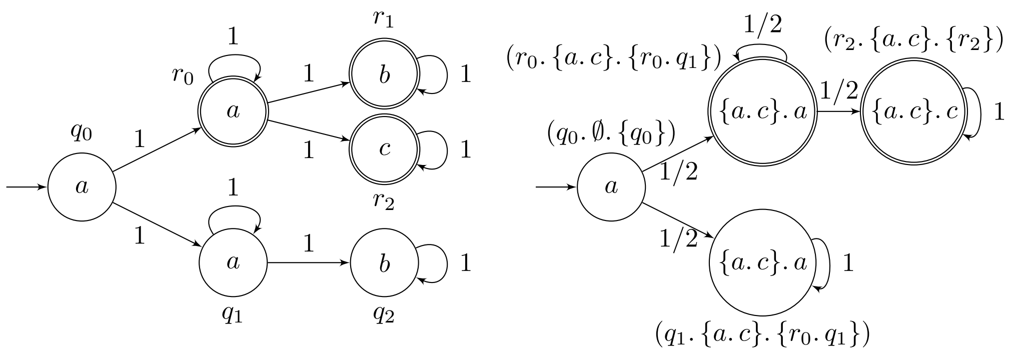

Figure 2 A WMC (left) and the finite oMC (right) induced by this WMC and the strategy that always allow $\{a, c\}$.

Writing $\mathcal{V}_{\mathbb{M}_\sigma(\mu_0)}$ for the set of infinite paths corresponding to AA-disclosing observation sequences in $\mathbb{M}_\sigma(\mu_0)$, we have $Disc^{\mathsf{AA}}(\mathbb{M}_\sigma(\mu_0)) = \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\mathcal{V}_{\mathbb{M}_\sigma(\mu_0)})$. Remark that an observation sequence of the oMC induced by a WMC and a strategy contains both the observation of the state of the WMC and the choices of allowed events done by the strategy. The observation sequence of a path in the induced oMC is therefore equal to the knowledge sequence of the corresponding path in the WMC. This choice of observation was done to express that the choices made by the strategy are known to the observer. An important consequence of this decision is that the strategy does not modify which observation sequences are AA-disclosing.

▶ **Lemma 14.** *Given $\mathbb{M}$ a WMC, μ_0 an initial distribution, $\mathsf{S}^{\mathsf{r}} \subseteq S$, σ, σ' two strategies and w an observation sequence produced by at least one path of $\mathbb{M}_\sigma(\mu_0)$ and one path of $\mathbb{M}_{\sigma'}(\mu_0)$, then $\mathsf{P^{rel}}_{\mathbb{M}_{\sigma'}(\mu_0)}(w) = \mathsf{P^{rel}}_{\mathbb{M}_\sigma(\mu_0)}(w)$.*

We study the two following diagnosability problems over WMC:

- The AA-*diagnosability problem* consists in, given a WMC $\mathbb{M}$ and an initial distribution μ_0, deciding whether there exists a strategy σ such that $\mathbb{M}_\sigma(\mu_0)$ is AA-diagnosable.
- The *strategy problem* consists in, given an AA-diagnosable WMC $\mathbb{M}$ with initial distribution μ_0, computing the strategy σ achieving $Disc^{\mathsf{AA}}(\mathbb{M}_\sigma(\mu_0)) = \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\mathsf{Rel})$.

▶ Remark 15. Even if this is not the usual framework of opacity, one may wonder whether it is possible to decide whether there exists a strategy allowing to obtain a disclosure above a given threshold. This can however easily be reduced to the emptiness problem of probabilistic automata which is well known to be undecidable [22]. Moreover, this reduction holds for all three notions of disclosure. This undecidability result vindicates the need for a specific, simpler framework for opacity.

▶ **Example 16.** Consider the WMC on the left of Figure 2. Without any control (*i.e.* with a strategy permanently allowing every event), one obtains the oMC of Figure 1, which is not AA-diagnosable. However, assuming 'b' is a controllable event, the strategy that always forbid it induces the oMC on the right of Figure 2 which is AA-diagnosable: every relevant path almost surely contains a 'c' that can not be generated by a non-relevant path. This oMC is in fact diagnosable exactly as once a 'c' occurs the proportion of relevant paths is equal to 1.

3.2 Solving AA-diagnosability for WMCs

While approximate diagnosability is simpler than exact diagnosability for oMC (PTIME vs PSPACE)[8, 7], for WMCs this difference disappears and both are EXPTIME-complete. The EXPTIME-completeness of exact diagnosis for WMC was established in [6]. We will devote this section to the proof of the following theorem:

▶ **Theorem 17.** *The* AA*-diagnosability is* EXPTIME*-complete.*

First, the hardness is established by a reduction from safety games with imperfect information [10]. This result is obtained directly by applying the proof of Proposition 3 of [6].

▶ **Proposition 18.** *The* AA*-diagnosability is EXPTIME-hard.*

Proof. In the proof of Proposition 3 of [6], a given safety game with imperfect information has a winning strategy iff no path is relevant. Moreover if a path is relevant, then its observation sequence is not AA-disclosing. More precisely, for ρ a relevant path, the proportion of relevant paths with observation sequence $K(\rho)$ is equal to $\frac{1}{2}$. Therefore the existence of a winning strategy in this game is equivalent to AA-diagnosability ensuring the EXPTIME-hardness. ◀

The most important step to solve AA-diagnosability for WMC is to develop a good understanding on the strategies optimising AA-disclosure. For starters, with a straightforward adaptation of a proof of [16], we show that one can consider deterministic strategies only.

▶ **Lemma 19.** *Given $\mathbb{M}$ a WMC, μ_0 an initial distribution, $\mathsf{S}^{\mathsf{r}} \subseteq S$ and σ a strategy, there exists a deterministic strategy σ' such that $Disc^{\mathsf{AA}}(\mathbb{M}_\sigma(\mu_0)) = \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\mathsf{Rel})$ implies $Disc^{\mathsf{AA}}(\mathbb{M}_{\sigma'}(\mu_0)) = \mathbf{P}_{\mathbb{M}_{\sigma'}(\mu_0)}(\mathsf{Rel})$.*

We can further restrict the strategies by limiting ourselves to belief-based strategy. This is far from an intuitive result. Indeed, while the AA-diagnosability of an oMC depends heavily on the exact values of the probabilities in the oMC, this result implies that the control only needs to remember the structure of the WMC. Remark though that the choice made in each belief depends on the probabilities.

▶ **Lemma 20.** *Given $\mathbb{M}$ a WMC, μ_0 an initial distribution, $\mathsf{S}^{\mathsf{r}} \subseteq S$ and σ a deterministic strategy, there exists a deterministic belief based strategy σ' such that $Disc^{\mathsf{AA}}(\mathbb{M}_\sigma(\mu_0)) = \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\mathsf{Rel})$ implies $Disc^{\mathsf{AA}}(\mathbb{M}_{\sigma'}(\mu_0)) = \mathbf{P}_{\mathbb{M}_{\sigma'}(\mu_0)}(\mathsf{Rel})$.*

Proof. Let $\mathbb{M}$ be a WMC, μ_0 be an initial distribution and σ be a deterministic strategy such that $\mathbb{M}_\sigma(\mu_0)$ is AA-diagnosable. We define a belief based strategy σ' from σ in the following way. Let $\rho \in \mathsf{FPath}(\mathbb{M}_\sigma(\mu_0))$. We define by E_ρ the set of finite path producing the same belief as ρ, *i.e.* $E_\rho = \{\rho' \in \mathsf{FPath}(\mathbb{M}_\sigma(\mu_0)) \mid B_{\mathbb{M}(\mu_0)}(\mathsf{O}(\rho')) = B_{\mathbb{M}(\mu_0)}(\mathsf{O}(\rho))\}$. We define $\sigma'(B_{\mathbb{M}(\mu_0)}(\mathsf{O}(\rho))) = \bigcup_{\rho' \in E_\rho} \sigma(\mathsf{O}(\rho'))$. Let us show that $\mathbb{M}_{\sigma'}(\mu_0)$ is AA-diagnosable.

Let two states $q = (s, \Sigma^\bullet, B) \in \mathsf{S}^{\mathsf{r}}$ and $q' = (s', \Sigma^\bullet, B) \in S \setminus \mathsf{S}^{\mathsf{r}}$ belonging to a BSCC of $\mathbb{M}_{\sigma'}(\mu_0)$ and reached by two finite paths ρ and ρ' of $\mathsf{FPath}(\mathbb{M}_{\sigma'}(\mu_0))$ with $\mathsf{O}(\rho) = \mathsf{O}(\rho')$. We will show that $d(\mathbb{M}_{\sigma'}(\mathbf{1}_q), \mathbb{M}_{\sigma'}(\mathbf{1}_{q'})) = 1$ using the characterisation given in Proposition 9. More precisely, for any observations sequence $w \in \Sigma^*$, and any pair of distributions on the set of states reached from q and from q' after observing w, we consider the probabilistic language generated by similar distributions in $\mathbb{M}_\sigma$ (*i.e.* distributions giving the same weight to the states of the original WMC $\mathbb{M}$) and rely on the fact that $\mathbb{M}_\sigma$ is AA-diagnosable to show that the generated languages are different. This implies the distance is 1 thanks to Proposition 9.

Let $w \in \Sigma^*$ such that $\mathbb{P}_{\mathbb{M}_{\sigma'(\mathbf{1}_q)}}(w) > 0$ and $\mathbb{P}_{\mathbb{M}_{\sigma'(\mathbf{1}_{q'})}}(w) > 0$, we denote by B_w, B_q and $B_{q'}$ the beliefs reached after observing w from the beliefs B, $\{q\}$ and $\{q'\}$ respectively, let two distributions μ_1' and μ_2' such that $\mathsf{Supp}(\mu_1') \subseteq B_q$, $\mathsf{Supp}(\mu_2') \subseteq B_{q'}$. As σ' does not allow events that are never allowed by σ in the same belief, there exists an observation sequence $w_\sigma \in \Sigma^*$ such that $\mathbb{P}_{\mathbb{M}_\sigma(\mu_0)}(w_\sigma) > 0$ and the belief reached in $\mathbb{M}(\mu_0)$ after a path of observation w_σ from the initial distribution is B_w, *i.e.* $B_{\mathbb{M}(\mu_0)}(w_\sigma) = B_w$. We can thus define initial distributions μ_1 and μ_2 on the set of states reached after observing w_σ in $\mathbb{M}_\sigma$ mimicking the distributions μ_1' and μ_2' (*i.e.* giving the same probability to configurations associated to the same state of $\mathbb{M}$). From the remark following Theorem 10 and Proposition 9, there exists a word w_d such that $\mathbb{P}_{\mathbb{M}_\sigma(\mu_1)}(w_d) \neq \mathbb{P}_{\mathbb{M}_\sigma(\mu_2)}(w_d)$. This implies that there exists a word w_d' such that $\mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(w_d') \neq \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_2')}(w_d')$. Indeed, let E be the set of observation sequences of the form $w'a$ where w' is a strict prefix of w_d, $a \in \Sigma$, $\mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(w'a) > 0$ and $\mathbb{P}_{\mathbb{M}_\sigma(\mu_1)}(w'a) = 0$. If $\mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(E) \neq \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_2')}(E)$, this implies our result. Else, by construction of the strategy σ' we have:

$$
\begin{aligned}
\mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(w_d) =& \mathbb{P}_{\mathbb{M}_\sigma(\mu_1)}(w_d) \times (1 - \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(E)) \\
\neq& \mathbb{P}_{\mathbb{M}_\sigma(\mu_2)}(w_d) \times (1 - \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_1')}(E)) \\
=& \mathbb{P}_{\mathbb{M}_\sigma(\mu_2)}(w_d) \times (1 - \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_2')}(E)) \\
=& \mathbb{P}_{\mathbb{M}_{\sigma'}(\mu_2')}(w_d),
\end{aligned}
$$

in which case we can choose $w_d' = w_d$. As this holds for any $w \in \Sigma^*$ and pair of distributions μ_1' and μ_2', according to Proposition 9 we have $d(\mathbb{M}_{\sigma'}(\mathbf{1}_q), \mathbb{M}_{\sigma'}(\mathbf{1}_{q'})) = 1$. From Theorem 10, we can thus deduce that $\mathbb{M}_{\sigma'}(\mu_0)$ is AA-diagnosable. Therefore belief-based strategies are sufficient to decide AA-diagnosability. ◀

A naive NEXPTIME algorithm can be obtained from these two lemmas: we guess a deterministic belief-based strategy then verify AA-diagnosability of the exponential oMC generated by the WMC and the strategy. In the following proposition, we show how to efficiently build a good belief-based strategy, which gives us an EXPTIME algorithm.

▶ **Proposition 21.** *The* AA*-diagnosability problem is in* EXPTIME*.*

Proof. Let $\mathbb{M}$ be a WMC and μ_0 be an initial distribution. To obtain the result, we first show that within a BSCC, the least restrictive a strategy is, the better it is for the purpose of diagnosis. However, a strategy too permissive may lead to the creation of new BSCC which may not be AA-diagnosable. Therefore, we will build a good strategy by an iterative procedure starting from the strategy allowing everything, then restricting it at each step to remove problematic BSCCs.

Let σ and σ' be two deterministic belief-based strategies such that for any belief B of $\mathbb{M}$ $\sigma(B) \subseteq \sigma'(B)$, let q be a relevant state associated to the belief B and belonging to a BSCC of both $\mathbb{M}_\sigma(\mu_0)$ and $\mathbb{M}_{\sigma'}(\mu_0)$. Assume that there exists a positive measure of paths in $\mathbb{M}_{\sigma'}(\mu_0)$ that visit q and that are not associated to an AA-disclosing observation sequence. Defining $B' = (B \setminus \mathsf{S}^\mathsf{r}) \cup \{q\}$, this is equivalent to saying that the WMC $\mathbb{M}_{\sigma'}(\mu_1)$, where μ_1 is an initial distribution of support B', is not AA-diagnosable. Therefore we can use the characterisation of Theorem 10. Without loss of generality, as q belongs to a BSCC, we can assume the pair of state given by the characterisation is (q, q') where $q' \notin \mathsf{S}^\mathsf{r}$, is associated to the belief B, belongs to a BSCC of $\mathbb{M}_{\sigma'}(\mu_1)$ and is such that $d(\mathbb{M}_{\sigma'}(\mathbf{1}_q), \mathbb{M}_{\sigma'}(\mathbf{1}_{q'})) < 1$. Let w, π_1 and π_2 be the observation sequence and the two distributions obtained by applying Proposition 9 on the pair of WMC $(\mathbb{M}_{\sigma'}(\mathbf{1}_q), \mathbb{M}_{\sigma'}(\mathbf{1}_{q'}))$. Let $q'' \notin \mathsf{S}^\mathsf{r}$ be a state belonging to a BSCC of $\mathbb{M}_\sigma(\mu_1)$ reachable from q' by a path which observation sequence w' is prefixed by w. Let π_1' and π_2' be the distribution obtained after observing w' starting in π_1 and π_2. As $\forall v \in \Sigma^*, \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_1)}(v) = \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_2)}(v)$, we also have $\forall v \in \Sigma^*, \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_1')}(v) = \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_2')}(v)$. This implies that $\forall v \in \Sigma^*, \mathbf{P}_{\mathbb{M}_{\sigma}(\pi_1')}(v) = \mathbf{P}_{\mathbb{M}_{\sigma}(\pi_2')}(v)$. Indeed, given $v \in \Sigma^*$, we have

$$
\begin{aligned}
\mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_1')}(v) &= \sum_{\rho \in \mathsf{O}^{-1}(v)} \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_1')}(\rho) \\
&= \sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(v)} \pi_1'(s_0) \prod_{i=0}^{n-1} \sigma'(K(v_{\downarrow 2i+1}))(\Sigma_i) p(s_{i+1} \mid s_i, \Sigma_i) \\
&= \prod_{i=0}^{n-1} \sigma'(K(v_{\downarrow 2i+1}))(\Sigma_i) \sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(v)} \pi_1'(s_0) \prod_{i=0}^{n-1} \frac{T(s_i, s_{i+1})}{\sum_{s'', \mathsf{O}(s'') \in \Sigma_i} T(s_i, s'')} \\
&= \prod_{i=0}^{n-1} \sigma'(K(v_{\downarrow 2i+1}))(\Sigma_i) \sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(v)} \pi_2'(s_0) \prod_{i=0}^{n-1} \frac{T(s_i, s_{i+1})}{\sum_{s'', \mathsf{O}(s'') \in \Sigma_i} T(s_i, s'')} \\
&= \mathbf{P}_{\mathbb{M}_{\sigma'}(\pi_2')}(v).
\end{aligned}
$$

As a consequence, $d(\mathbb{M}_\sigma(\mathbf{1}_q), \mathbb{M}_\sigma(\mathbf{1}_{q'})) < 1$. From the remark following Theorem 10, this implies that $\mathbb{M}_\sigma(\mu_1)$ is not AA-diagnosable and thus there exists a positive measure of paths in $\mathbb{M}_\sigma(\mu_0)$ that visit q and that are not associated to an AA-disclosing observation sequence. Therefore, having restricted the strategy σ' did not allow to regain AA-diagnosability of the paths visiting q. This means that a strategy achieving AA-diagnosability of the WMC must ensure that q cannot be reached.

Using this result, we build iteratively the most permissive strategy ensuring AA-diagnosability. We start with the strategy σ_0 allowing everything. Assume we built the strategy σ_k such that any less permissive strategy do not ensure AA-diagnosability. If $\mathbb{M}_{\sigma_k}(\mu_0)$ is not AA-diagnosable, there exists two states s and s' associated to the same belief B that satisfies the characterisation of Theorem 10. W.l.o.g one can assume that both of these states belong to BSCCs of $\mathbb{M}_{\sigma_k}(\mu_0)$. From our preliminary result, we know that any strategy that contains the states s and s' in a BSCC does not ensure AA-diagnosability. As any strategy less permissive than σ_k does not ensure AA-diagnosability, we need to restrict the strategy so that s and s' are not reachable, or that s and s' are not in BSCCs anymore. The latter is in fact not sufficient as the remark following Theorem 10 would still apply on this pair of states. Thus we build σ_{k+1} as the most permissive strategy such that $\mathbb{M}_{\sigma_{k+1}}(\mu_0)$ does not contain s and s'. This can easily be done by belief based strategies as removing these states is equivalent to removing the belief B. This procedure ends when the strategy σ_n that is created either is the most permissive strategy ensuring AA-diagnosability or if one cannot build a strategy removing the problematic states/belief. This algorithm is in EXPTIME as every step of the procedure can be done in exponential time (verification of AA-diagnosability, identification of the pair of problematic states and creation of the new strategy are all steps that can be done in EXPTIME) and there is at most exponentially many steps as each one of them removes at least one belief from the system, and there are exponentially many beliefs. Therefore, the AA-diagnosability problem can be solved in EXPTIME. ◀

The previous proof building the strategy ensuring AA-diagnosability when it exists, this algorithm also solves the strategy problem.

4 Active Approximate Opacity

As discussed at the beginning of Section 3, the framework of the study of active opacity is different from the one used for active diagnosis. While most elements are similar, strategies are given more power in the way they observe and affect the system. Moreover, the goal of the strategies is now either to maximise or to minimise the disclosure of information depending on whether they are deemed adversarial or cooperative.

4.1 Opacity for Observable Markov Decision Processes

▶ **Definition 22** (oMDP). *An observable Markov Decision Process (oMDP) over alphabet Σ is a tuple $\mathsf{M} = (S, \mathsf{Act}, p, \mathsf{O})$ where S is a finite set of states, $\mathsf{Act} = \cup_{s \in S} A(s)$ where $A(s)$ is a finite non-empty set of actions for each state $s \in S$, $p : S \times \mathsf{Act} \to \mathsf{Dist(S)}$ is the (partial) transition function defined for (s, a) when $a \in A(s)$ and $\mathsf{O} : S \to \Sigma$ is the observation function.*

As before, we write $p(s'|s, a)$ instead of $p(s, a)(s')$. Given an initial distribution μ_0, an infinite path of $\mathsf{M}(\mu_0)$ is a sequence $\rho = s_0 a_0 s_1 a_1 \ldots$ where $\mu_0(s_0) > 0$ and $p(s_{i+1}|s_i, a_i) > 0$, for $s_i \in S$, $a_i \in A(s_i)$, for all $i \geq 0$. Finite paths are defined like for WMC, and we use similar notations for the various sets of paths. Given a path $\rho = s_0 a_0 s_1 a_1 \ldots s_i$ its observation is $\mathsf{O}(\rho) = \mathsf{O}(s_0)\mathsf{O}(s_1) \ldots \mathsf{O}(s_i)$.

The nondeterministic choice of the action is resolved by strategies.

▶ **Definition 23** (Strategy for oMDP). *A strategy for an oMDP M with initial distribution μ_0 is a mapping $\sigma : \mathsf{FPath}(\mathsf{M}(\mu_0)) \to \mathsf{Dist}(\mathsf{Act})$ associating with any finite path ρ a distribution $\sigma(\rho)$ on the actions in $A(\mathsf{last}(\rho))$.*

Recall that strategies for WMCs were making their choice based on the knowledge sequence alone. This represented that the strategy was extern and thus only had partial information on the system. The oMDP framework however gives to the strategy full knowledge of the path. Similarly as for WMC, given a strategy σ, a path $\rho = s_0 a_0 s_1 a_1 \ldots$ of M is σ-compatible if for all i, $a_i \in \mathsf{Supp}(\sigma(s_0 a_0 s_1 a_1 \ldots s_i))$. A strategy σ is *observation-based* if for any finite path ρ, $\sigma(\rho)$ only depends on the observation $O(\rho)$ and on the last state $\mathsf{last}(\rho)$. We can also adapt the notions of deterministic and belief-based strategies.

A strategy σ on $\mathsf{M}(\mu_0)$ defines a (possibly infinite) oMC $\mathsf{M}_\sigma(\mu_0)$ with set of states $\mathsf{FPath}(\mathsf{M}_\sigma(\mu_0))$ (the finite σ-compatible paths), that can be equipped with the observation function associating $\mathsf{O}(\mathsf{last}(\rho))$ with the finite path ρ. The transition function p_σ is defined for $\rho \in \mathsf{FPath}(\mathsf{M}_\sigma(\mu_0))$ and $\rho' = \rho a s'$ by $p_\sigma(\rho'|\rho) = \sigma(\rho)(a) p(s'|s, a)$ and we denote by $\mathbf{P}_{\mathsf{M}_\sigma(\mu_0)}$ the associated probability measure. The definition of the observation function shows that the observer of the system does not know what action is chosen by the strategy at any step. However, the observer still knows which strategy was selected initially, allowing him to deduce the oMC M_σ.

Disclosure values for oMDP are defined according to the status of the strategies, by considering them as adversarial or cooperative with respect to the system.

▶ **Definition 24** (Disclosure of an oMDP). *Given an oMDP $\mathsf{M} = (S, \mathsf{Act}, p, \mathsf{O})$, an initial distribution μ_0 and a set of relevant states $\mathsf{S^r} \subseteq S$, the maximal AA-disclosure of $\mathsf{S^r}$ in $\mathsf{M}(\mu_0)$ is $Disc^{\mathsf{AA}}_{\max}(\mathsf{M}(\mu_0)) = \sup_\sigma Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0))$ and the minimal AA-disclosure is $Disc^{\mathsf{AA}}_{\min}(\mathsf{M}(\mu_0)) = \inf_\sigma Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0))$.*

We study the following opacity problems over oMDP:

- **Quantitative decision problems:** The *minimal* AA-*disclosure problem* consists in, given an oMDP M and a threshold $\theta \in [0,1]$, deciding if $Disc^{\mathsf{AA}}_{min}(\mathsf{M}) \leq \theta$? The *maximal* AA-*disclosure problem* consists in, given an oMDP M and a threshold $\theta \in [0,1]$, deciding if $Disc^{\mathsf{AA}}_{max}(\mathsf{M}) \geq \theta$?
- **Qualitative decision problems:** The *limit-sure disclosure problem* is the special case of the AA-disclosure problem with $\theta = 1$ for maximisation and with $\theta = 0$ for minimisation and the *almost-sure disclosure problem* consists in deciding whether there exists a strategy achieving a disclosure of 1 for maximisation and 0 for minimisation.

4.2 Possible Restriction on the Strategies

The decidability result for AA-diagnosability of WMC relied strongly on a restriction to a sufficient subset of strategies. It is thus natural to take a similar approach for opacity. We can indeed establish some such restriction, for instance to observation-based strategies. This is proven for an exact notion of disclosure in [2], however the very same proof applies to the accurate approximate notion.

▶ **Proposition 25** ([2]). *Given an oMDP, an initial distribution μ_0, $\mathsf{S}^{\mathsf{r}} \subseteq S$ and a strategy σ, there exists an observation-based strategy σ' such that $Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) = Disc^{\mathsf{AA}}(\mathsf{M}_{\sigma'}(\mu_0))$.*

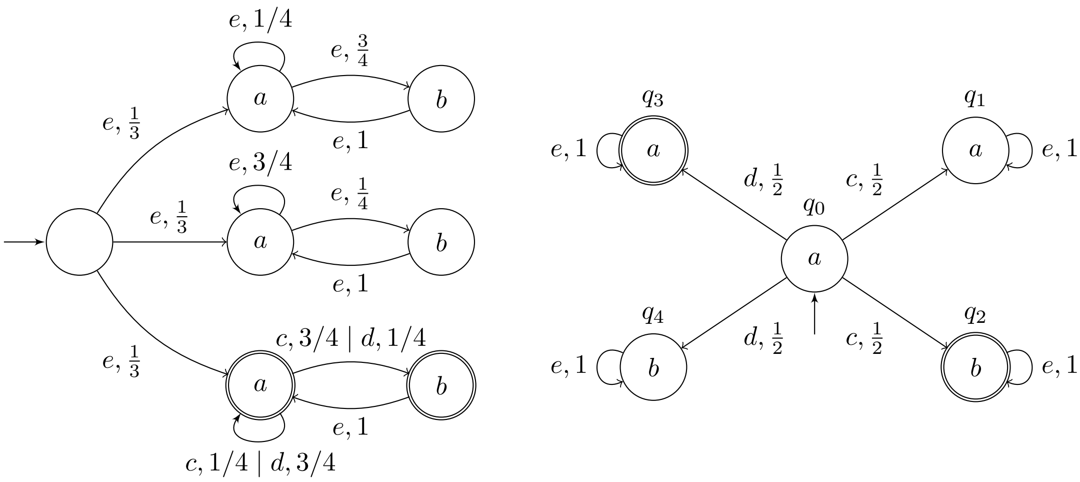

Figure 3 Left: An oMDP where randomisation or non-belief based strategies are necessary to maximise the AA-disclosure. Right: An oMDP where randomisation is necessary to minimise the disclosure.

However, the restriction cannot be extended to deterministic belief-based strategies. Consider the example on the left of Figure 3 with maximisation of the AA-disclosure in mind. There are three components, in each of them a 'b' is always followed by an 'a', however, the probability that a 'b' occurs after an 'a' varies. This probability is $\frac{3}{4}$ in the upper component, $\frac{1}{4}$ in the middle one and depends on the strategy on the one below. A deterministic belief based strategy will either always choose the action 'c' or always the action 'd'. Such a strategy replicates the probabilistic behaviour of one of the other two components, inducing an AA-disclosure of 0. However a randomised strategy, giving for instance a half probability to both actions obtains a $\frac{1}{2}$ probability to produce a 'b' after an 'a'. This belief-based randomised strategy induces then an AA-disclosure of $\frac{1}{3}$. One could define a deterministic strategy which is not belief based and obtain a disclosure of $\frac{1}{3}$ too by alternating the choices of the action $'c'$ and $'d'$. Therefore, maximising strategies require randomisation, more memory than just the belief or both[2].

When aiming to minimise the AA-disclosure, we can show that randomisation is necessary. Consider the oMDP depicted on the right of Figure 3. The strategy only has to make a choice between two actions during the first step. Thus, there are only two existing deterministic strategies, choosing respectively 'c' or 'd' in q_0. In both cases, the disclosure is $\frac{1}{2}$. On the other hand, any randomized strategies σ_p such that $\sigma_p(q_0)$ activates 'c' with probability p and 'd' with probability $(1-p)$ with $0 < p < 1$, induces an oMC that do not contain any AA-disclosing observation, hence the disclosure is 0.

4.3 (Un)decidability of the Opacity Problems

The examples of the previous subsection point to the idea that the traditional framework for active opacity is more complicated than the one considered for active diagnosability. This is confirmed by the (un)decidability results that we establish below. The undecidability proofs we establish are done by reduction of problems in probabilistic automata.

Let us first consider the maximisation of AA-disclosure.

▶ **Theorem 26.** *The maximal* AA*-disclosure problem is undecidable. The maximal limit-sure disclosure problem is undecidable.*

As a silver lining, the almost-sure AA-disclosure problem is easily decidable.

▶ **Theorem 27.** *The maximal almost-sure* AA*-disclosure problem is in PTIME.*

These results are not exactly surprising as opacity problems for maximisation had already been shown to be undecidable for exact notions of opacity in [2]. However, while in this same paper the authors show that most opacity problems for minimisation are decidable, these problems become undecidable for the accurate approximate notion of opacity.

▶ **Theorem 28.** *The minimal almost-sure and the minimal limit-sure* AA*-disclosure decision problems are undecidable.*

▶ **Corollary 29.** *The minimal* AA*-disclosure decision problem is undecidable.*

References

1 B. Bérard, K. Chatterjee, and N. Sznajder. Probabilistic opacity for Markov decision processes. *Information Processing Letters*, 115(1):52–59, 2015.

2 B. Bérard, S. Haddad, and E. Lefaucheux. Probabilistic disclosure: Maximisation vs. minimisation. In *Proceedings of FSTTCS'17*, volume 93 of *LIPIcs*, pages 13:1–13:14. Leibniz-Zentrum für Informatik, 2017.

3 B. Bérard, O. Kouchnarenko, J. Mullins, and M. Sassolas. Preserving opacity on interval Markov chains under simulation. In *Proceedings of WODES'16*, pages 319–324. IEEE, 2016.

4 B. Bérard, O. Kouchnarenko, J. Mullins, and M. Sassolas. Opacity for linear constraint Markov chains. *Discrete Evente Dynamic Systems*, 28(1):83–108, 2018.

5 B. Bérard, J. Mullins, and M. Sassolas. Quantifying opacity. *Mathematical Structures in Computer Science*, 25(2):361–403, 2015.

[2] Using a complicated example, one can show that randomisation cannot always substitute the need for additional memory. This also holds for minimising strategies.

6 N. Bertrand, E. Fabre, S. Haar, S. Haddad, and L. Hélouët. Active diagnosis for probabilistic systems. In *Proceedings of FoSSaCS'14*, volume 8412 of *LNCS*, pages 29–42. Springer, 2014.

7 N. Bertrand, S. Haddad, and E. Lefaucheux. Foundation of diagnosis and predictability in probabilistic systems. In *Proceedings of FSTTCS'14*, volume 29 of *LIPIcs*, pages 417–429. Leibniz-Zentrum für Informatik, 2014.

8 N. Bertrand, S. Haddad, and E. Lefaucheux. Accurate approximate diagnosability of stochastic systems. In *Proceedings of LATA'16*, volume 9618 of *LNCS*, pages 549–561. Springer, 2016.

9 Nathalie Bertrand, Serge Haddad, and Engel Lefaucheux. A Tale of Two Diagnoses in Probabilistic Systems. *Information and Computation*, page 104441, 2019.

10 D. Berwanger and L. Doyen. On the power of imperfect information. In *Proceedings of FSTTCS'08*, volume 2 of *LIPIcs*, pages 73–82. Leibniz-Zentrum für Informatik, 2008.

11 A. Borodin. On relating time and space to size and depth. *SIAM J. Comput.*, 6:733–744, 12 1977.

12 A. Borodin, J. [von zur Gathen], and J. Hopcroft. Fast parallel matrix and gcd computations. *Information and Control*, 52(3):241 – 256, 1982.

13 J. W. Bryans, M. Koutny, L. Mazaré, and P. Y. A. Ryan. Opacity generalised to transition systems. *Intl. Journal of Information Security*, 7(6):421–435, 2008.

14 F. Cassez and S. Tripakis. Fault diagnosis with static and dynamic observers. *Fundamenta Informaticae*, 88:497–540, 2008.

15 E. Chanthery and Y. Pencolé. Monitoring and active diagnosis for discrete-event systems. *IFAC Proceedings Volumes*, 42(8):1545 – 1550, 2009.

16 K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *Proceedings of MFCS'10*, volume 6281 of *LNCS*, pages 246–257. Springer, 2010.

17 T. Chen and S. Kiefer. On the total variation distance of labelled Markov chains. In *Proceedings of CSL-LICS'14*, pages 33:1–33:10. ACM, 2014.

18 H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *Proceedings of ICALP'10*, volume 6199 of *LNCS*, pages 527–538. Springer, 2010.

19 S. Jiang, Z. Huang, V. Chandra, and R. Kumar. A polynomial algorithm for testing diagnosability of discrete-event systems. *Transactions on Automatic Control*, 46(8):1318–1321, 2001.

20 A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *SWAT'72*, pages 125–129. IEEE, 1972.

21 K Mulmuley. A fast parallel algorithm to compute the rank of a matrix over an arbitrary field. In *STOC '86*, page 338–339, 1986.

22 A. Paz. *Introduction to probabilistic automata*. Academic Press, 1971.

23 A. Saboori and Ch. N. Hadjicostis. Current-state opacity formulations in probabilistic finite automata. *Transactions on Automatic Control*, 59(1):120–133, 2014.

24 M. Sampath, S. Lafortune, and D. Teneketzis. Active diagnosis of discrete-event systems. *Transactions on Automatic Control*, 43(7):908–929, 1998.

25 M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *Transactions on Automatic Control*, 40(9):1555–1575, 1995.

26 D. Thorsley and D. Teneketzis. Diagnosability of stochastic discrete-event systems. *Transactions on Automatic Control*, 50(4):476–492, 2005.

27 D. Thorsley and D. Teneketzis. Active acquisition of information for diagnosis and supervisory control of discrete-event systems. *Discrete Event Dynamic Systems*, 17:531–583, 2007.

A Equivalence of the AA-diagnosability definitions

▶ **Proposition 8.** *An oMC is* AA*-diagnosable iff it is* ε*-diagnosable for all* $\varepsilon > 0$.

Proof. Let $\mathcal{M}$ be a finite oMC and μ_0 an initial distribution.

Suppose that $\mathcal{M}(\mu_0)$ is AA-diagnosable. By definition, given an AA-disclosing observation sequence w, for all $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $w_{\downarrow n}$ is ε-disclosing. Therefore for all $\varepsilon > 0$, $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) \leq Disc^{\varepsilon}(\mathcal{M}(\mu_0))$. Moreover, as $\mathcal{M}$ is AA-diagnosable, $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = \mathbf{P}(\mathsf{Rel})$. Thus, $Disc^{\varepsilon}(\mathcal{M}(\mu_0)) \geq \mathbf{P}(\mathsf{Rel})$. Finally, by definition of $Disc^{\varepsilon}$, for all $\varepsilon > 0$ $Disc^{\varepsilon}(\mathcal{M}(\mu_0)) \leq \mathbf{P}(\mathsf{Rel})$. Thus $Disc^{\varepsilon}(\mathcal{M}(\mu_0)) = \mathbf{P}(\mathsf{Rel})$ and $\mathcal{M}(\mu_0)$ is ε-diagnosable.

Conversely, suppose that $\mathcal{M}(\mu_0)$ is not AA-diagnosable. Let us consider the set of infinite words $D = \cap_{\varepsilon>0} D^{\varepsilon}_{min}\Sigma^{\omega} \setminus D^{\mathsf{AA}}$. Let us show that $\mathbf{P}(D) = 0$. Let $w \in D$, we have (1) for all $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that ${\mathsf{P}^{\mathsf{rel}}}_{\mathcal{M}(\mu_0)}(w_{\downarrow n}) > 1 - \varepsilon$ and (2) $({\mathsf{P}^{\mathsf{rel}}}_{\mathcal{M}(\mu_0)}(w_{\downarrow n}))_{n\in\mathbb{N}}$ does not converge toward 1. Given $\varepsilon > 0$ and denoting by E_ε the set of ε-min-disclosing observaion sequence, due to (1) we have

$$\begin{aligned}
\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(D) \setminus \mathsf{Rel}\}) &< \sum_{w\in E_\varepsilon} \mathbf{P}(\{\rho \in \mathsf{O}^{-1}(w) \setminus \mathsf{Rel}\})\\
&< \sum_{w\in E_\varepsilon} \mathbf{P}(\{\rho \in \mathsf{O}^{-1}(w) \cap \mathsf{Rel}\})\frac{\varepsilon}{1-\varepsilon}\\
&< \frac{\varepsilon}{1-\varepsilon}.
\end{aligned}$$

As this holds for all $\varepsilon > 0$, $\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(D) \setminus \mathsf{Rel}\}) = 0$. Moreover, due to (2), there exists $\varepsilon > 0$ such that for infinitely many $n \in \mathbb{N}$ we have ${\mathsf{P}^{\mathsf{rel}}}_{\mathcal{M}(\mu_0)}(w_{\downarrow n}) < 1 - \varepsilon$. For all $k \in \mathbb{N}$, we denote by E_k the set of prefixes w of words of D such that ${\mathsf{P}^{\mathsf{rel}}}_{\mathcal{M}(\mu_0)}(w) < 1 - \varepsilon$ for the k'th time. We then have for all k:

$$\begin{aligned}
\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(E_k) \setminus \mathsf{Rel}\}) &= \sum_{w\in E_k} \mathbf{P}(\{\rho \in \mathsf{O}^{-1}(w) \setminus \mathsf{Rel}\})\\
&> \sum_{w\in E_k} \mathbf{P}(\{\rho \in \mathsf{O}^{-1}(w) \cap \mathsf{Rel}\})\frac{\varepsilon}{1-\varepsilon}\\
&> \frac{\varepsilon}{1-\varepsilon}\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(D) \cap \mathsf{Rel}\})
\end{aligned}$$

As $(\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(E_k) \setminus \mathsf{Rel}\}))_{k\in\mathbb{N}}$ converges toward $\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(D) \setminus \mathsf{Rel}\})$ which is equal to 0, this implies that $\mathbf{P}(\{\rho \in \mathsf{O}^{-1}(D) \cap \mathsf{Rel}\}) = 0$ and thus that $\mathbf{P}(D) = 0$. As a consequence, $\lim_{\varepsilon\to 0} \mathbf{P}(D^{\varepsilon}_{min}) = \mathbf{P}(D^{\mathsf{AA}})$. As $\mathcal{M}(\mu_0)$ is not AA-diagnosable by assumption, there thus exists $\varepsilon > 0$ such that $\mathcal{M}(\mu_0)$ is not ε-diagnosable. ◀

B AA-disclosure problem for oMC

▶ **Theorem 11.** *The* AA*-disclosure problem for finite oMC is* PSPACE*-complete.*

Proof. Let us first show how to solve the AA-disclosure problem in EXPTIME. We first build an exponential size oMC which contains additional information compared to the original one. Then we show that there are two kinds of BSCC in this new oMC: the ones that are reached by paths that almost surely have an AA-disclosing observation sequence, and the ones that are reached by paths that do not correspond to AA-disclosing observation sequences. We can then use the existing results for the AA-diagnosability problem to determine the status of each BSCC. Therefore, computing the AA-disclosure of the oMC is equivalent to computing the probability to reach the "AA-disclosing" BSCC, which can be done in NC in the size of the oMC, thus giving an overall PSPACE algorithm.

Let $\mathcal{M} = (S, p, \mathsf{O})$ be a finite oMC and μ_0 be an initial distribution. We build a new oMC $\mathcal{M}' = (S', p', \mathsf{O}')$ which has the same behaviour as $\mathcal{M}$ but where the states are enriched with an additional information (the set of states the system can be in, given the produced observation sequence):

- $S' = S \times 2^S$;
- For $(s,B),(s',B') \in S'$, $p'((s',B') \mid (s,B)) = p(s' \mid s)$ if $B' = \cup_{q\in B}\mathsf{Supp}(p(q)) \cap \mathsf{O}^{-1}(\mathsf{O}(s'))$ else, $p'((s',B') \mid (s,B)) = 0$;
- For $(s,B) \in S'$, $\mathsf{O}'(s,B) = \mathsf{O}(s)$.

We define the initial distribution μ'_0 for $\mathcal{M}'$ by $\mu'_0(s, \mathsf{Supp}(\mu_0) \cap \mathsf{O}^{-1}(\mathsf{O}(s))) = \mu_0(s)$ for all $s \in S$. There is a one-to-one correspondence between the paths of $\mathcal{M}(\mu_0)$ and $\mathcal{M}'(\mu'_0)$: every path $\rho = s_0 s_1 \cdots s_n$ of $\mathcal{M}(\mu_0)$ is associated to an unique path $\rho' = (s_0,B_0)(s_1,B_1)\cdots(s_n,B_n)$ with $\mathsf{O}(\rho) = \mathsf{O}(\rho')$, $\mathbf{P}_{\mathcal{M}(\mu_0)}(\rho) = \mathbf{P}_{\mathcal{M}'(\mu'_0)}(\rho')$ and B_n contains the set of states of S that can be reached with a path of observation $\mathsf{O}(\rho)$. Due to the latter property, B_n only depends on $\mathsf{O}(\rho)$ and is called the *belief* associated to $\mathsf{O}(\rho)$.

Let $(s,B) \in S'$ such that $s \in \mathsf{S}^{\mathsf{r}}$ and (s,B) belongs to a BSCC of $\mathcal{M}'$.We claim that either for every path ρ ending in (s,B), $\mathbf{P}(\{\rho' \in \mathsf{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge \mathsf{O}(\rho') \in D^{\mathsf{AA}}\}) = 0$ or for every path ρ ending in (s,B), $\mathbf{P}(\{\rho' \in \mathsf{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge \mathsf{O}(\rho') \in D^{\mathsf{AA}}\}) = \mathbf{P}(\rho)$. In other words, there are two categories of BSCC composed of relevant states: the ones that almost surely accurate approximately disclose the relevance and the ones that do not accurate approximately disclose the relevance at all. Moreover, the BSCC containing (s,B) do not disclose the relevance at all iff there exists a state $s' \in B$ such that s' belongs to a BSCC of B, $s' \notin \mathsf{S}^{\mathsf{r}}$ and $d(\mathcal{M}(\mathbf{1}_s), \mathcal{M}(\mathbf{1}_{s'})) < 1$. The proof of this claim can be obtained in a straightforward manner from the proof of Theorem 10. For the sake of pedagogy, we give some elements of this proof below.

Assume that for all $s' \in B$ such that s' belongs to a BSCC of B and $s' \notin \mathsf{S}^{\mathsf{r}}$ we have $d(\mathcal{M}(\mathbf{1}_s), \mathcal{M}(\mathbf{1}_{s'})) = 1$. Then denoting $B' = (B \setminus \mathsf{S}^{\mathsf{r}}) \cup \{s\}$, then Lemma B of [9] directly tells us that for any initial distribution μ_1 of support B', we have that $\mathcal{M}'(\mu_1)$ is AA-diagnosable. As the states of $B \setminus B'$ can only increase the relevance proportion, this ensures that $\mathbf{P}(\{\rho' \in \mathsf{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge \mathsf{O}(\rho') \in D^{\mathsf{AA}}\}) = \mathbf{P}(\rho)$.

Conversely, if there exists a state $s' \in B$ such that s' belongs to a BSCC of B, $s' \notin \mathsf{S}^{\mathsf{r}}$ and $d(\mathcal{M}(\mathbf{1}_s), \mathcal{M}(\mathbf{1}_{s'})) < 1$, then one can rely on the proof of Lemma A of [9] to obtain the result. We develop the proof here in the simpler case where B does not contain any relevant state beside s. Using Proposition 9 and the correspondence between $\mathcal{M}$ and $\mathcal{M}'$, one deduces that there exists $\rho_{(s,B)} \in \mathsf{FPath}(\mathcal{M}(\mathbf{1}_{(s,B)}))$ and $\alpha > 0$ such that for all $w \in \Sigma^*$ with $\mathsf{O}(\rho) \leq w$

$$\begin{aligned}
&\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(\{\rho' \in \mathsf{FPath}(\mathcal{M}'(\mathbf{1}_{(s,B)})) \mid \rho_{(s,B)} \preceq \rho' \wedge \mathsf{O}(\rho') = w\}) \\
&\qquad \leq \alpha \mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s',B)})}(\{\rho' \in \mathsf{FPath}(\mathcal{M}'(\mathbf{1}_{(s',B)})) \mid \mathsf{O}(\rho') = w\}).
\end{aligned}$$

Therefore, for all $w \in \Sigma^*$ and initial distribution μ_1 of support B we have:

$$\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}'(\mu_1)}(w) \leq \frac{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(w)}{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(w) + \frac{\mu_1(s')}{\mu_1(s)}\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s',B)})}(w)} \tag{1}$$

$$\leq \frac{\varepsilon_w + \displaystyle\sum_{\rho \mid \mathsf{O}(\rho\rho_{(s,B)}) \leq w} \frac{\alpha \mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(\rho)}{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(\rho_{(s,B)})} \mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s',B)})}(w^\rho)}{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(w) + \frac{\mu_1(s')}{\mu_1(s)}\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s',B)})}(w)} \tag{2}$$

where $\varepsilon_w = \mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(\{\rho \in \mathsf{FPath}(\mathcal{M}(\mathbf{1}_{(s,B)}) \mid \nexists \rho_1, \rho_2, \rho = \rho_1 \rho_{(s,B)} \rho_2 \wedge \mathsf{O}(\rho) = w\})$ and w^ρ is such that $w = \mathsf{O}(\rho) w^\rho$. As with probability 1, a run of $\mathcal{M}'(\mathbf{1}_{(s,B)})$ visits (s,B) infinitely often, it will almost surely contain a $\rho_{(s,B)}$ subrun, more precisely: the value $\frac{\varepsilon_w}{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(w)}$ almost surely converges to 0 when $|w|$ diverges to ∞. Let $w \in \Sigma^\omega$, if $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}'(\mu_1)}(w_{\downarrow n}) \xrightarrow{n \to \infty} 1$

then, for all ρ such that $\mathsf{O}(\rho\rho_{(s,B)}) \leq w$ we have that $\frac{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s',B)})}(w^{\rho}_{\downarrow n})}{\mathbf{P}_{\mathcal{M}'(\mathbf{1}_{(s,B)})}(w_{\downarrow n})}$ converges to 0, thus, due to Equation 2, $\varepsilon_{w_{\downarrow n}}$ does not converge to 0, which can only happen with probability 0. Therefore $\mathsf{P^{rel}}_{\mathcal{M}'(\mu_1)}(w_{\downarrow n})$ almost surely does not converge to 1. This implies that $\mathbf{P}\{\rho' \in \mathsf{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge \mathsf{O}(\rho') \in D^{AA}\} = 0$.

This result establishes that the BSCC of $\mathcal{M}'$ are partitionned between the ones that accurately approximately and almost surely disclose the relevance and the ones that do not accurately approximately disclose it at all. Moreover, one can detect in PTIME (in the size of the original oMC $\mathcal{M}$) what kind of BSCC a given state belongs to. Therefore, one can obtain the value of $Disc^{AA}(\mathcal{M}'(\mu'_0))$ by computing the probability to reach the disclosing BSCC, which is known to be possible in PTIME in the size of $\mathcal{M}'$. In fact, as computing this probability amount to solve a linear system of equations, this can even be done in NC [12, 21]. The oMC $\mathcal{M}'$ being exponential in the size of $\mathcal{M}$, and as NC blown up to the exponential is equal to PSPACE [11], this yields a PSPACE algorithm. As $Disc^{AA}(\mathcal{M}(\mu_0)) = Disc^{AA}(\mathcal{M}'(\mu'_0))$, this allows us to solve the AA-disclosure problem.

We now establish the hardness by reducing the universality problem for non-deterministic finite automaton (NFA), which is known to be PSPACE-complete [20].

Let $\mathcal{A} = (Q, \Sigma, T, q_0, F)$ be an NFA (Q is the set of states, q_0 the initial one, F the set of accepting states, Σ the alphabet and $T \in Q \times \Sigma \times Q$ the transition function). W.l.o.g. we can assume that $F = Q$ and $\Sigma = \{a, b\}$. Our first step is to push the observations onto the states (as shown in Figure 4). From $\mathcal{A}$ we define the incomplete oMC $\hat{\mathcal{A}} = (S_A, p_A, O_A)$ and the initial distribution μ^A_0 such that:

- $S_A = Q \times \Sigma$;
- for $(q,c),(q',d) \in S_A$, if $(q,d,q') \in T$, then $p_A((q',d) \mid (q,c)) = \frac{1}{|S_A|+1}$, else $p_A((q',d) \mid (q,c)) = 0$;
- for $(q,c) \in S_A, O_A(q,c) = c$;
- for $(q',d) \in S_A$, if $(q_0,d,q') \in T$, then $\mu^A_0(q',d) = \frac{1}{|S_A|+1}$, else $\mu^A_0(q',d) = 0$.

This oMC is incomplete as none of the distributions μ^A_0 and $p_A(\cdot \mid s)$ (for $s \in S_A$) sum to 1.

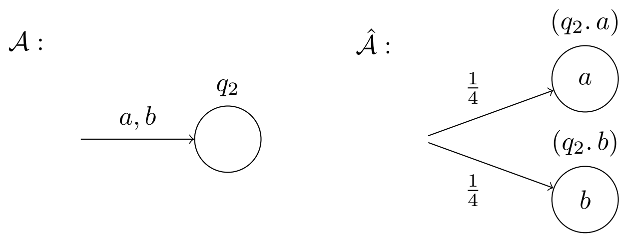

Figure 4 From NFA $\mathcal{A}$ to incomplete oMC $\hat{\mathcal{A}}$. The label next to the state is its name. We will not always display the state's name so as not to overload the figure.

We now build the oMC $\mathcal{M} = (S, p, O)$ represented in Figure 5 where

- $S = S_A \cup \{s_\sharp, f_a, f_b, f_\sharp\}$;
- given $s, s' \in S_A$, $p(s' \mid s) = p_A(s' \mid s)$, $p(s_\sharp \mid s) = 1 - \sum_{s' \in S_A} p(s' \mid s)$, for $h \in \{f_a, f_b\}$ and $g \in \{f_a, f_b, f_\sharp\}$, $p(g \mid h) = 1/3$ and $p(f_\sharp \mid f_\sharp) = p(s_\sharp \mid s_\sharp) = 1$;
- for $s \in S_A$, $O(s) = O_A(s)$, $O(s_\sharp) = O(f_\sharp) = \sharp, O(f_a) = a$ and $O(f_b) = b$.

We also define μ_0 as $\mu_0(s) = \mu^A_0(s)$ for $s \in S_A$ and $\mu_0(f_a) = \mu_0(f_b) = \frac{1 - \sum_{s \in S_A} \mu_0(s)}{2}$.

Choosing $\mathsf{S^r} = \{f_a, f_b, f_\sharp\}$, let us show that $\mathcal{A}$ is not universal iff $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) > 0$.

Suppose first that $\mathcal{A}$ is not universal. There thus exists a word $w \in \Sigma*$ such that no path starting in S_A has observation sequence w. As there exists one relevant path

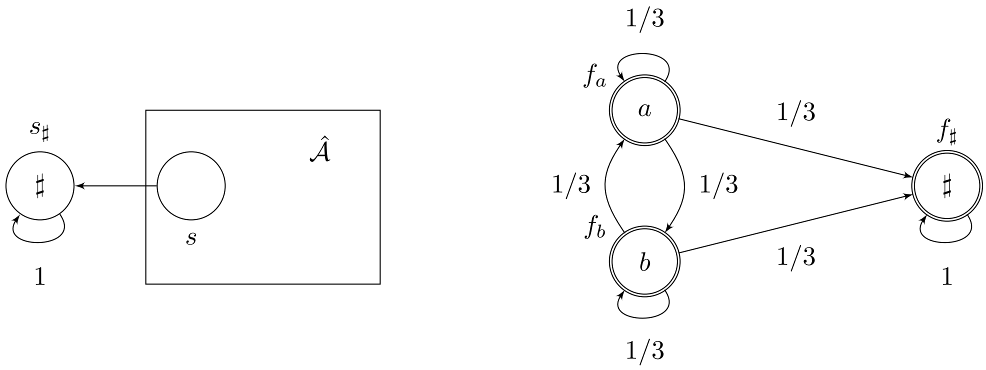

Figure 5 A reduction for PSPACE-hardness of the AA-disclosure problem.

ρ (starting in either f_a or f_b) associated to $w\sharp$, we have $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w\sharp) = 1$. Therefore $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) \geq \mathbf{P}_{\mathcal{M}(\mu_0)}(\rho) > 0$.

Conversely, assume that $\mathcal{A}$ is universal. Let ρ be a path ending in $f_\sharp$ with observation sequence $\mathsf{O}(\rho) = w\sharp$ for some $w \in \Sigma^*$. As $\mathcal{A}$ is universal, there exists a finite path ρ' in $\hat{\mathcal{A}}$ with observation sequence w. As for every state s of $\hat{\mathcal{A}}$, $p(s_\sharp \mid s) > 0$, ρ' can be extended into a finite path ρ'' ending in $s_\sharp$ with observation $w\sharp$. Thus, $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w\sharp) < 1$. Moreover, every path ending with a $\sharp$ remains with probability 1 in either $s_\sharp$ or $f_\sharp$, due to this for every $k \geq 2$, $\mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w\sharp^k) = \mathsf{P}^{\mathsf{rel}}{}_{\mathcal{M}(\mu_0)}(w\sharp)$. Therefore, $w\sharp^\omega \notin D^{\mathsf{AA}}$. This implies that no infinite path visiting $f_\sharp$ corresponds to an AA-disclosing observation sequence. $f_\sharp$ being the only relevant state, $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = 0$. ◀

C The AA-disclosing observation sequence do not depend on the strategy in WMCs

▶ **Lemma 14.** *Given $\mathbb{M}$ a WMC, μ_0 an initial distribution, $\mathsf{S}^{\mathsf{r}} \subseteq S$, σ, σ' two strategies and w an observation sequence produced by at least one path of $\mathbb{M}_\sigma(\mu_0)$ and one path of $\mathbb{M}_{\sigma'}(\mu_0)$, then $\mathsf{P}^{\mathsf{rel}}{}_{\mathbb{M}_{\sigma'}(\mu_0)}(w) = \mathsf{P}^{\mathsf{rel}}{}_{\mathbb{M}_\sigma(\mu_0)}(w)$.*

Proof. Let $\mathbb{M}$ be a WMC, μ_0 be an initial distribution, σ be a strategy and $w = o_0\Sigma_0 \dots \Sigma_{n-1}o_n$ be an observation sequence produced by at least one path of $\mathbb{M}_\sigma(\mu_0)$. By definition of w, $\mathsf{P}^{\mathsf{rel}}{}_{\mathbb{M}_\sigma(\mu_0)}(w)$ is defined and in particular $\prod_{i=0}^{n-1} \sigma(K(w_{\downarrow 2i+1}))(\Sigma_i) \neq 0$. We have

$$\begin{aligned}
\mathsf{P}^{\mathsf{rel}}{}_{\mathbb{M}_\sigma(\mu_0)}(w) &= \frac{\mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\{\rho \in \mathsf{O}^{-1}(w) \mid \rho \in \mathsf{Rel}\})}{\mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(w)} \\
&= \frac{\sum_{\rho \in \mathsf{O}^{-1}(w) \mid \rho \in \mathsf{Rel}} \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\rho)}{\sum_{\rho \in \mathsf{O}^{-1}(w)} \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\rho)} \\
&= \frac{\sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(w) \mid \rho \in \mathsf{Rel}} \prod_{i=0}^{n-1} \sigma(K(w_{\downarrow 2i+1}))(\Sigma_i) p(s_{i+1} \mid s_i, \Sigma_i)}{\sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(w)} \prod_{i=0}^{n-1} \sigma(K(w_{\downarrow 2i+1}))(\Sigma_i) p(s_{i+1} \mid s_i, \Sigma_i)} \\
&= \frac{\sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(w) \mid \rho \in \mathsf{Rel}} \prod_{i=0}^{n-1} p(s_{i+1} \mid s_i, \Sigma_i)}{\sum_{\rho = s_0\Sigma_0 \dots s_n \in \mathsf{O}^{-1}(w)} \prod_{i=0}^{n-1} p(s_{i+1} \mid s_i, \Sigma_i)}
\end{aligned}$$

which is independent of σ, therefore for any strategy σ' such that at least one path of $\mathbb{M}_{\sigma'}(\mu_0)$ produces w, $\mathsf{P^{rel}}_{\mathbb{M}_{\sigma'}(\mu_0)}(w) = \mathsf{P^{rel}}_{\mathbb{M}_{\sigma}(\mu_0)}(w)$. ◀

D Deterministic strategies for AA-diagnosability

▶ **Lemma 19.** *Given $\mathbb{M}$ a WMC, μ_0 an initial distribution, $\mathsf{S^r} \subseteq S$ and σ a strategy, there exists a deterministic strategy σ' such that $Disc^{\mathsf{AA}}(\mathbb{M}_\sigma(\mu_0)) = \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(\mathsf{Rel})$ implies $Disc^{\mathsf{AA}}(\mathbb{M}_{\sigma'}(\mu_0)) = \mathbf{P}_{\mathbb{M}_{\sigma'}(\mu_0)}(\mathsf{Rel})$.*

Proof. In the proof of Lemma 1 of [16], the authors show that a randomised 'observation based' strategy can be seen as an average over a family of deterministic 'observation based' strategies[3]. A consequence of their equation (2) in our framework is the following: given a strategy σ, for every set of path E, there exists a deterministic strategy σ_{det} such that (a) $\mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0)) \subseteq \mathsf{Path}(\mathbb{M}_\sigma(\mu_0))$ and (b) $\mathbf{P}_{\mathbb{M}_{\sigma_{det}}(\mu_0)}(E) \geq \mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(E)$. Using this result with the appropriate set E we will show that if $\mathbb{M}_\sigma(\mu_0)$ is AA-diagnosable then $\mathbb{M}_{\sigma_{det}}(\mu_0)$ is AA-diagnosable.

We define $E_\sigma = \mathcal{V}_{\mathbb{M}_\sigma(\mu_0)} \cup (\mathsf{Path}(\mathbb{M}_\sigma(\mu_0)) \setminus \mathsf{Rel})$ which are the set of σ-compatible paths that are either not relevant or AA-disclosing. Let σ_{det} be the strategy obtained by applying the result of [16] on the set E_σ. Suppose $\mathbb{M}_\sigma(\mu_0)$ is AA-diagnosable. By definition, this is equivalent to $\mathbf{P}_{\mathbb{M}_\sigma(\mu_0)}(E_\sigma) = 1$. Due to (b), this implies that $\mathbf{P}_{\mathbb{M}_{\sigma_{det}}(\mu_0)}(E_\sigma) = 1$ too. Moreover $\mathcal{V}_{\mathbb{M}_{\sigma_{det}}(\mu_0)} = \mathcal{V}_{\mathbb{M}_\sigma(\mu_0)} \cap \mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0))$, thanks to Lemma 14 and (a). Thus

$$\begin{aligned} E_\sigma &= \mathcal{V}_{\mathbb{M}_{\sigma_{det}}(\mu_0)} \cup (\mathcal{V}_{\mathbb{M}_\sigma(\mu_0)} \setminus \mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0)) \cup (\mathsf{Path}(\mathbb{M}_\sigma(\mu_0)) \setminus \mathsf{Rel}) \\ &= E_{\sigma_{det}} \cup (\mathcal{V}_{\mathbb{M}_\sigma(\mu_0)} \cup (\mathsf{Path}(\mathbb{M}_\sigma(\mu_0)) \setminus \mathsf{Rel}) \setminus \mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0)) \end{aligned}$$

where $E_{\sigma_{det}} = \mathcal{V}_{\mathbb{M}_{\sigma_{det}}(\mu_0)} \cup (\mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0)) \setminus \mathsf{Rel})$.

Finally, $\mathbf{P}_{\mathbb{M}_{\sigma_{det}}(\mu_0)}(\mathcal{V}_{\mathbb{M}_\sigma(\mu_0)} \cup (\mathsf{Path}(\mathbb{M}_\sigma(\mu_0)) \setminus \mathsf{Rel}) \setminus \mathsf{Path}(\mathbb{M}_{\sigma_{det}}(\mu_0)) = 0$ as no path of this set is σ_{det}-compatible. Therefore $\mathbf{P}_{\mathbb{M}_{\sigma_{det}}(\mu_0)}(E_{\sigma_{det}}) = 1$ which implies that $\mathbb{M}_{\sigma_{det}}(\mu_0)$ is AA-diagnosable. ◀

E Maximisation of the AA-disclosure

Recall first that a probabilistic automata (PA) is a tuple $\mathfrak{A} = (Q, q_0, \Sigma, T, F)$ where Q is a finite set of states with $q_0 \in Q$ the initial state, Σ is a finite alphabet (which cumulates the role of the actions in the oMDP and of the observation), $T : Q \times \Sigma \to \mathsf{Dist(Q)}$ is the transition function and $F \subseteq Q$ is the set of final states. We define paths for PA as usual and for a finite path $\rho = q_0 a_1 q_1 \ldots a_n q_n$ of $\mathfrak{A}$, the word $a_1 \ldots a_n \in \Sigma^*$ is called the *trace* of ρ and denoted by $tr(\rho)$. Writing $\mathsf{FPath}_{(w,q)}(\mathfrak{A}) = \{\rho \in \mathsf{FPath}(\mathfrak{A}) \mid tr(\rho) = w \text{ and } \mathsf{last}(\rho) = q\}$ for $w \in \Sigma^*$ and $q \in Q$, we define $\mathbf{P}_\mathfrak{A}(w, q) = \mathbf{P}_\mathfrak{A}(\cup_{\rho \in \mathsf{FPath}_{(w,q)}(\mathcal{A})} \mathsf{Cyl}(\rho))$, $\mathbf{P}^F_\mathfrak{A}(w, F) = \sum_{q \in F} \mathbf{P}_\mathcal{A}(w, q)$ and $\mathsf{Val}(\mathfrak{A}) = \sup_{w \in \Sigma^*} \mathbf{P}_\mathcal{A}(w, F)$.

Given a threshold $\theta \in (0, 1)$, we set $\mathcal{L}_{>\theta}(\mathfrak{A}) = \{w \in \Sigma^* \mid \mathbf{P}_\mathfrak{A}(w, F) > \theta\}$. The strict emptiness problem for $\mathfrak{A}$ consists in asking whether $\mathcal{L}_{>\theta}$ is empty, and is known to be undecidable for $\theta > 0$ [22]. The value 1 problem, *i.e.* asking whether $val(\mathfrak{A}) = 1$, is undecidable as well [18].

▶ **Theorem 26.** *The maximal* AA*-disclosure problem is undecidable. The maximal limit-sure disclosure problem is undecidable.*

[3] In our framework, by definition, every strategy is 'observation based'.

Proof. These results are obtained by reductions from the strict emptiness problem and value 1 problem on probabilistic automata. As a consequence, we first need a method to adapt a given probabilistic automaton to our framework. This transformation bears many similarities with what was done for NFA in the beginning of Theorem 11. For simplicity, we use states without observations (denoted by the observation ε), this is without loss of generality as we could remove them using a simple probabilistic closure since no non-deterministic choice occurs within them.

Given a probabilistic automaton $\mathfrak{A} = (Q, q_0, \{a,b\}, T, F)$ over $\{a,b\}$ that we suppose complete (*i.e.* $T(q,c)$ is defined for all $q \in Q$ and $c \in \{a,b\}$) without loss of generality, we first transform $\mathfrak{A}$ into an incomplete oMDP $\hat{\mathfrak{A}} = (\hat{Q}, \{e\}, \hat{p}, \hat{\mathsf{O}})$ over the observation alphabet $\{a,b\}$ where the observations are pushed from the transitions to the next state (an illustration is given in Figure 6). The set of states is $\hat{Q} = Q \cup \{q_c \mid q \in Q \wedge c \in \{a,b\}\}$. The observation function $\hat{\mathsf{O}}$ is defined by $\hat{\mathsf{O}}(q) = \varepsilon$ and $\hat{\mathsf{O}}(q_c) = c$ for $q \in Q$ and $c \in \{a,b\}$. The transition function $\hat{p}$ is defined for $q, q' \in Q$ and $c \in \{a,b\}$ by $\hat{p}(q' \mid q_c, e) = T(q' \mid q, c)$ and $\hat{p}(q_c \mid q, e) = \frac{1}{4}$. This oMDP is incomplete as the probabilities do not sum to 1. Intuitively, a letter to read is chosen at random, and then the transition is taken according to the probabilities induced by the chosen letter. Remark that the strategy do not make any choice here.

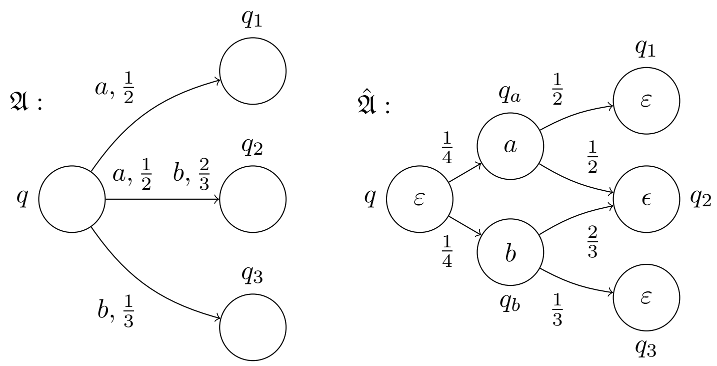

Figure 6 From PA $\mathfrak{A}$ to incomplete oMDP $\hat{\mathfrak{A}}$. The action e labelling each transition is omitted in the oMDP.

From $\hat{\mathfrak{A}}$ we build the oMDP $\mathsf{M} = (S, \{e,r\}, p, \mathsf{O})$ (represented in Figure 7) where:

- $S = \hat{Q} \cup \{s_0, s^u_f, s^u_n, s^u_s\} \cup \{s^z_t \mid z \in \{m,l\}, t \in \{\varepsilon, a, b, \sharp, \flat\}\}$;
- $p(q_0 \mid s_0, e) = p(q^m_\varepsilon \mid s_0, e) = 1/2$, for $q, q' \in \hat{Q}, p(q' \mid q, e) = \hat{p}(q' \mid q, e)$, for $q \in F, q' \in Q \setminus F, p(s^u_n \mid q, e) = p(s^u_f \mid q', e) = 1/2$, $p(q_0 \mid s^u_f, e) = 1, p(q_0 \mid s^u_n, e) = p(s^u_s \mid s^u_n, e) = 1/2$, for $z \in \{m,l\}, c, c' \in \{a,b\}, p(s^z_c \mid s^z_\varepsilon, e) = 1/2, p(s^z_{c'} \mid s^z_c, e) = 1/4, p(s^z_\sharp \mid s^z_c, e) = 1/2$, $p(s^m_\flat \mid s^m_\sharp, r) = p(s^m_\varepsilon \mid s^m_\sharp, r) = 1/2, p(s^l_\varepsilon \mid s^m_\sharp, e) = 1$ and $p(s^l_\flat \mid s^l_\sharp, e) = p(s^l_\varepsilon \mid s^l_\sharp, e) = 1/2$;
- for $q \in \hat{Q}, \mathsf{O}(q) = \hat{\mathsf{O}}(q)$, $\mathsf{O}(s_0) = \mathsf{O}(s^u_f) = \mathsf{O}(s^u_n) = \sharp$ $\mathsf{O}(s^u_s) = \flat$ and for $z \in \{m,l\}, t \in \{\varepsilon, a, b, \sharp, \flat\}$ $\mathsf{O}(s^z_t) = t$.

The set of relevant states is defined as $\mathsf{S}^\mathsf{r} = \{s^u_s\}$ and we consider the initial distribution $\mu_0 = \mathbf{1}_{q_0}$.

We will show that, for a given threshold λ, there exists a strategy σ such that $Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) > \lambda/2$ iff there exists a word $w \in \{a,b\}^*$ such that $\mathbf{P}_\mathcal{A}(w, F) > \lambda$.

Let us first give the intuition behind this construction. The MDP M is composed of three parts (upper, middle and lower part of the Figure 7). The upper part mostly imitates the behaviour of the PA $\mathfrak{A}$ on random words, a $\sharp$ signalling the end of the word. If the run is

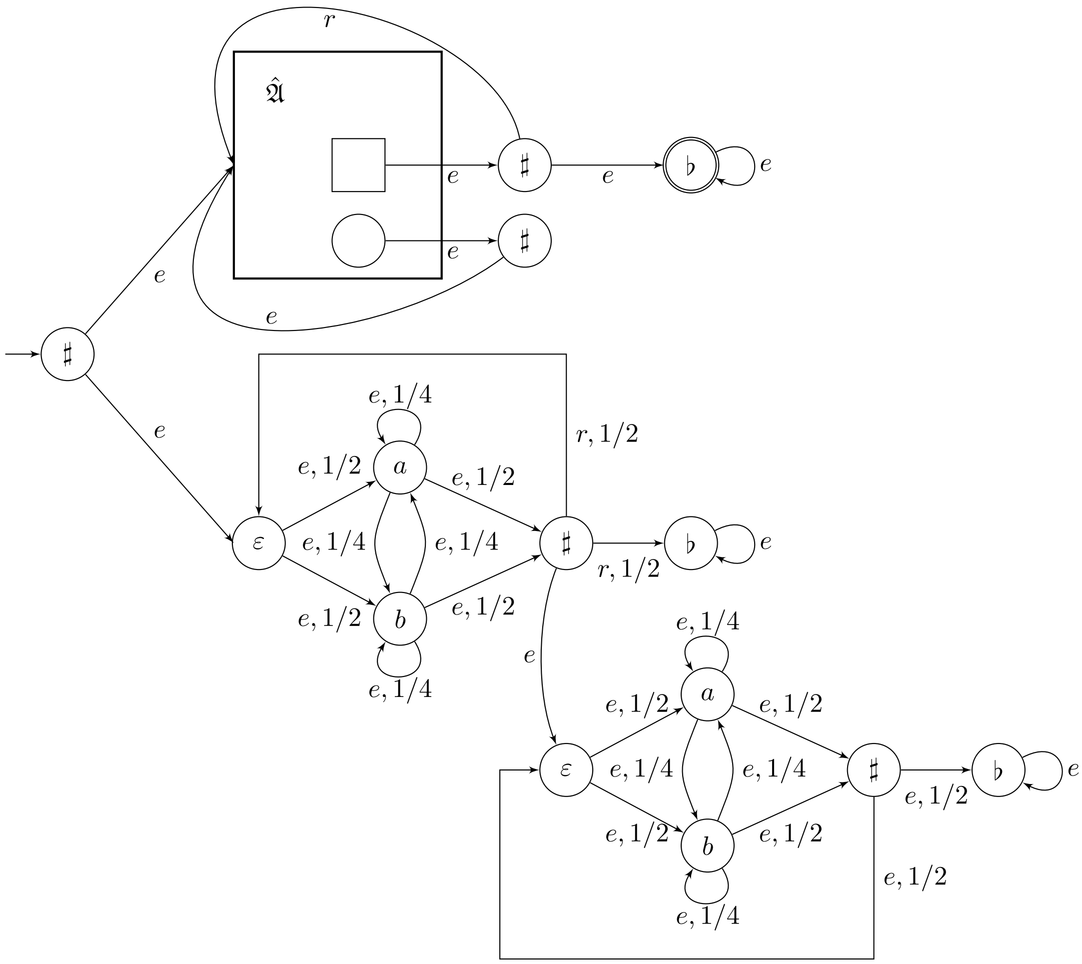

Figure 7 Reduction from the emptiness problem to the maximal disclosure problem. The square state corresponds to a final state of $\mathfrak{A}$.

accepting, *i.e.* if it ended in a final state of tha PA, then the strategy may chose to play e in order to reach the secret state, otherwise a new word is read. As the strategy knows in which state the system is, it could ‘cheat’ and reach the secret almost surely in the upper part. However, the middle and lower parts are used to make this additional knowledge of the strategy useless: after reading a word w in the middle part, the strategy chooses between the action e and r, using the action r implies that $w\sharp\flat^\omega$ is not AA-disclosing while using the action e makes $w\sharp\flat^\omega$ AA-disclosing but the run also reaches the lower part ensuring that any other observation from then on is not AA-disclosing. In other words, the strategy will have to choose a set of words for which it plays e simultaneously in the middle and the upper parts.

Formally, let us first identify which relevant paths are disclosing with a strategy σ. Let ρ be a relevant path with observation $w\flat^k$ for some $k \in \mathbb{N}$ and $w \in \{a, b, \sharp\}^*$. As once reaching a state labelled by $\flat$, there is no probabilistic behaviour, ${\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}_\sigma(\mu_0)}(w\flat) = {\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}(\mu_0)}(w\flat^k)$. Thus, the infinite observation associated to the unique infinite path extending ρ is AA-disclosing iff ${\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}(\mu_0)}(w\flat) = 1$. This happens iff for every path ρ' such that $\mathsf{O}(\rho') = w$ and $\mathsf{last}(\rho') = s^m_\sharp$, $\sigma(\rho') = e$ and there does not exist any path ρ'' such that $\mathsf{O}(\rho'') = w$ and $\mathsf{last}(\rho'') = s^l_\sharp$. This last condition can also be formulated as the absence of any run which observation is a prefix of w, ending in $s^m_\sharp$ and for which σ selects the action e.

let $\lambda \in \mathbb{R}$ assume that there exists a word $w \in \{a, b\}^*$ such that $\mathbf{P}_\mathcal{A}(w, F) > \lambda$. We define the strategy σ such that given a path with observation $w_1\sharp w_2\sharp \dots w_k\sharp$ such that for all $i \le k, w_i \in \{a, b\}^*$, if $w_k = w$ and both e and r are allowed actions in $\mathsf{last}(\rho)$, then σ chooses e, otherwise it chooses r if possible. This strategy induces a disclosure greater than $\lambda/2$. Indeed, in the upper part of the oMDP, in between two occurrences of $\sharp$ there is a positive probability that w is observed. Thus, with probability 1 a word of the form $w_1\sharp w_2\sharp \dots w_k\sharp$ such that for all $i \le k, w_i \in \{a, b\}^*$, for all $i < k, w_i \neq w$ and $w_k = w$ will be triggered. Moreover, let v be one such word, then, thanks to the choice of the strategy and the remark of the previous paragraph, ${\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}_\sigma(\mu_0)}(v\flat) = 1$. Finally, the probability that a path of the upper part of the oMDP, with observation v ends in s^u_n (allowing to trigger $\flat$ on the next step) is $\mathbf{P}_\mathcal{A}(w, F)$, thus ensuring that $Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) > \lambda/2$.

Conversely, assume that there exists a strategy σ such that $Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) > \lambda/2$. We define the set of words $E = \{(w, w') \in \{a, b\} * \times \{a, b, \sharp\}^* \mid \exists u \in \{a, b, \sharp\}^*, w' = u\sharp w\sharp \wedge \mathsf{O}{-}1(w'\flat) \neq \emptyset \wedge {\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}(\mu_0)}(w'\flat) = 1\}$. Relying on the earlier remark on which paths are disclosing, we have

$$
\begin{aligned}
Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) &= \sum_{(w,w')\in E} \mathbf{P}_{\mathsf{M}_\sigma(\mu_0)}(\{\rho \in \mathsf{FPath}(\mathsf{M}_\sigma(\mu_0)) \mid \mathsf{O}(\rho) = w'\flat\}) \\
&\le \sum_{(w,w')\in E} \mathbf{P}_{\mathsf{M}_\sigma(\mu_0)}(\{\rho \in \mathsf{FPath}(\mathsf{M}_\sigma(\mu_0)) \mid \mathsf{O}(\rho) = w' \wedge \mathsf{last}(\rho) = s^u_n\}) \\
&= \sum_{(w,w')\in E} \mathbf{P}_{\mathsf{M}_\sigma(\mu_0)}(w') \cdot 1/2\mathbf{P}_\mathcal{A}(w, F) \\
&\le 1/2 \max_{(w,w')\in E} \mathbf{P}_\mathcal{A}(w, F)
\end{aligned}
$$

Therefore, $Disc^{\mathsf{AA}}(\mathsf{M}_\sigma(\mu_0)) > \lambda/2$ implies that there exists w such that $\mathbf{P}_\mathcal{A}(w, F) > \lambda$.

This equivalence directly shows that the maximal AA-disclosure problem is undecidable. For the maximal limit-sure disclosure, one can use the same reduction with one additional secret state with observation $\natural$ (thus disclosing) which is reached with positive probability from any state s^m_c with $c \in \{a, b\}$. This means that the longer we wait before selecting a word, the higher the probability that a path that went to the middle part is AA-disclosing. However, the remaining paths are enough to guarantee the same reasoning as before for the paths going to the upper part, thus showing undecidability of maximal limit-sure disclosure. ◀

▶ **Theorem 27.** *The maximal almost-sure* AA*-disclosure problem is in PTIME.*

Proof. Given $\mathcal{M}$ an oMC, let us show that $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = 1$ iff $\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{Rel}) = 1$.

First, $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) \leq \mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{Rel})$ by definition, thus if $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) = 1$ then $\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{Rel}) = 1$.

Conversely, suppose that $Disc^{\mathsf{AA}}(\mathcal{M}(\mu_0)) < 1$, there thus exists a set of infinite observations E_o that are not AA-disclosing and such that $\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{O}^{-1}(E_o) \cap \mathsf{Rel}) > 0$. By definition of AA-disclosing, there thus exists $\varepsilon > 0$ such that there exists a subset of E_o, denoted E_ε, of infinite observations for which none of their prefixes are ε-disclosing and $\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{O}^{-1}(E_\varepsilon) \cap \mathsf{Rel}) = \lambda > 0$. By definition of $\mathsf{P}^{\mathsf{rel}}$, this implies that $\mathbf{P}_{\mathcal{M}(\mu_0)}(\mathsf{O}^{-1}(E_\varepsilon) \setminus \mathsf{Rel}) > \frac{\lambda\varepsilon}{(1-\varepsilon)}$. Therefore, $\mathbf{P}(\mathsf{Rel}) < 1 - \frac{\lambda\varepsilon}{(1-\varepsilon)} < 1$.

Given M an oMDP, M is thus almost-surely AA-disclosing iff there exists a strategy σ such that $\mathbf{P}_{\mathsf{M}_\sigma(\mu_0)}(\mathsf{Rel}) = 1$. As Rel is defined by the reachability of a set of states, this is equivalent to almost-sure reachability in MDP which is known to be in PTIME. ◀

F Minimal disclosure

▶ **Theorem 28.** *The minimal almost-sure and the minimal limit-sure* AA*-disclosure decision problems are undecidable.*

Proof. Given a probabilistic automaton $\mathfrak{A} = (Q, \{a, b\}, q_0, T, F)$ over $\{a, b\}$ we first transform $\mathfrak{A}$ into an incomplete MDP $\hat{\mathcal{A}} = (\hat{Q}, \{e\}, \hat{p}, \hat{\mathsf{O}})$ as in the proof of Theorem 26.

From $\hat{\mathcal{A}}$ we build the MDP $\mathsf{M} = (S, \{e, c, l\}, p, \mathsf{O})$ (represented in Figure 8) where:

- $S = \hat{Q} \cup \{s_0, s_0^1, s_a^1, s_b^1, s_1^1, s_\flat^1, s_\sharp^1, s_f^2, s_u^2, s_\flat^2, s_\sharp^2\}$,
- $p(q_0 \mid s_0, e) = p(s_0^1 \mid s_0, e) = \frac{1}{2}$. For $q_1, q_2 \in \hat{Q}$, $p(q_2 \mid q_1, e) = \hat{p}(q_2 \mid q_1, e)$. For $q \in Q$, if $q \in F$ then $p(s_f^2 \mid q, e) = \frac{1}{2}$ else $p(s_u^2 \mid q, e) = \frac{1}{2}$. $p(s_a^1 \mid s_0^1, e) = p(s_b^1 \mid s_0^1, e) = 1/2, p(s_a^1 \mid s_a^1, e) = p(s_a^1 \mid s_b^1, e) = p(s_b^1 \mid s_a^1, e) = p(s_b^1 \mid s_b^1, e) = 1/4, p(s_1^1 \mid s_a^1, e) = p(s_1^1 \mid s_b^1, e) = 1/2$. $p(s_\flat^2 \mid s_f^2, e) = p(s_\sharp^2 \mid s_u^2, e) = 1$, $p(s_\sharp^1 \mid s_1^1, l) = 1$, $p(s_\sharp^1 \mid s_1^1, c) = p(s_f^1 lat \mid s_1^1, c) = \frac{1}{2}$. $p(q_0 \mid s_\sharp^2, e) = p(q_0 \mid s_\flat^2, e) = p(s_0^1 \mid s_\sharp^1, e) = p(s_0^1 \mid s_\flat^1, e) = 1$. Undefined transitions have value 0.
- $\mathsf{O}(q) = \hat{\mathsf{O}}(q)$ for $q \in \hat{Q}$, $\mathsf{O}(s_d^i) = d$ for $i \in \{1, 2\}$ and $d \in \{\flat, \sharp, a, b\}$, $\mathsf{O}(s_0) = \mathsf{O}(s_0^1) = \varepsilon$ and $\mathsf{O}(s) = \sharp$ otherwise.

We consider the initial distribution $\mu_0 = \mathbf{1}_{s_0}$ and the set of relevant states $\mathsf{S}^{\mathsf{r}} = \hat{Q} \cup \{s_f^2, s_u^2, s_\flat^2, s_\sharp^2\}$ (i.e. the upper component of the system). We will show that $Disc_{min}^{\mathsf{AA}}(\mathsf{M}) > 0$ iff there exists a word w such that $\mathbb{P}_{\mathfrak{A}}(w) > \frac{1}{2}$.

The idea of the proof is the following. During the first transition one goes with same probability in s_O^1 or q_0^2 (lower and upper systems of the Figure 8. Then, on both side a word $w \in (a + b)^*$ is read with same probability, a $\sharp$ marking the end of the word. On the upper side, this $\sharp$ is followed by a $\flat$ with probability $\mathbb{P}_{\mathfrak{A}}(w)$ and by a $\sharp$ otherwise. On the lower side, a $\flat$ is read with a probability chosen by the controller between 0 and $\frac{1}{2}$ and a $\sharp$ otherwise. Therefore, the controller can reproduce the same probability on both side of the system (and thus give no information to the observer) iff the acceptance probability of w in $\mathfrak{A}$ is between 0 and $\frac{1}{2}$. The execution then starts again from the initial state of both copies of the automaton.

More formally, suppose that there exists a word $w_d \in \{a, b\}^*$ such that $\mathbb{P}_{\mathfrak{A}}(w_d) > \frac{1}{2}$. Given a finite observation $w \in \Sigma^*$, we define the value $ratio_{w_d}(w)$ as the ratio between the number of occurrence of $(\sharp + \flat)w_d\sharp\flat$ over the number of occurrence of $(\sharp + \flat)w_d\sharp$ in $\sharp w$.

This definition is extended to infinite observations by taking the limit, when defined, of the ratios of its finite prefixes. Let σ be any strategy. We define the set of observations $E = \{w \in \Sigma^\omega \mid ratio_{w_d}(w) > 1/2\}$. Thanks to the weak law of large numbers and by choice of w_d, we have that $\mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0eq_0})} = 1$ and $\mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0es_0^1})} = 0$. As from the initial state, a path of $\mathsf{M}_\sigma(\mu_0)$ goes either in s_0eq_0, becoming a relevant path, or $s_0es_0^1$, from which it can never become relevant, this implies that with probability 1, a relevant path has an observation belonging to E. Let us show that these relevant paths are almost surely AA-disclosing which will imply that $Disc^{\mathsf{AA}}_{min}(\mathsf{M}(\mu_0)) = \frac{1}{2}$.

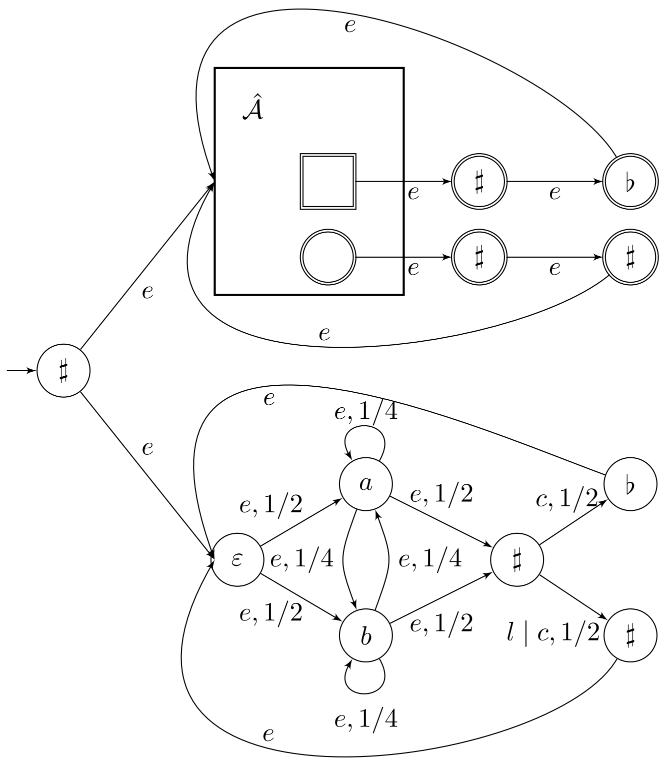

Figure 8 Reduction from the emptiness problem to the minimal almost-sure disclosure problem. The square state corresponds to a final state of $\mathfrak{A}$.

For every $n \in \mathbb{N}$, let $\mathfrak{S}_n$ be the set of prefixes of length n of the observations of E: $\mathfrak{S}_n = \{\sigma \in \Sigma_o^n \mid \exists \sigma' \in E, \sigma \preceq \sigma'\}$. For every $\varepsilon > 0$, we also define $\mathfrak{S}_n^\varepsilon$ as the subset of $\mathfrak{S}_n$ consisting of observations whose proportion of relevant paths exceeds threshold $1-\varepsilon$ in $\mathsf{M}(\mu_0)$: $\mathfrak{S}_n^\varepsilon = \{\sigma \in \mathfrak{S}_n \mid \mathsf{P^{rel}}_{\mathsf{M}(\mu_0)}(\sigma) < 1-\varepsilon\}$.

From $\bigcap_{n\in\mathbb{N}} \mathsf{Cyl}(\mathfrak{S}_n) = E$, we derive that $\lim_{n\to\infty} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0es_0^1})}(\mathfrak{S}_n) = \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0es_0^1})}(E) = 0$. Thus $\lim_{n\to\infty} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0es_0^1})}(\mathfrak{S}_n^\varepsilon) = 0$.

On the other hand, for every $n \in \mathbb{N}$,

$$\mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e s_0^1})}(\mathfrak{S}_n^\varepsilon) = \sum_{\sigma \in \mathfrak{S}_n^\varepsilon} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e s_0^1})}(\sigma) > \sum_{\sigma \in \mathfrak{S}_n^\varepsilon} \frac{\varepsilon}{1-\varepsilon} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e q_0})}(\sigma) = \frac{\varepsilon}{1-\varepsilon} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e q_0})}(\mathfrak{S}_n^\varepsilon) \ .$$

Since ε is fixed, $\mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e q_0})}(\mathfrak{S}_n^\varepsilon) < \frac{1-\varepsilon}{\varepsilon} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e s_0^1})}(\mathfrak{S}_n^\varepsilon)$ and $\lim_{n\to\infty} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e s_0^1})}(\mathfrak{S}_n^\varepsilon) = 0$ imply that $\lim_{n\to\infty} \mathbf{P}_{\mathsf{M}_\sigma(\mathbf{1}_{s_0 e q_0})}(\mathfrak{S}_n^\varepsilon) = 0$. This implies that with probability 1, a path whose observation belongs to E is ε-disclosing. As this holds for every $\varepsilon > 0$, from Proposition 8, we deduce that the infinite paths with observations in E are almost surely AA-disclosing.

Conversely, suppose that every word $w \in \{a,b\}^*$ verifies $\mathbb{P}_{\mathfrak{A}}(w) = \lambda \leq \frac{1}{2}$. We define the strategy σ such that after a path ρ ending in s_1^1 with an observation $\sharp w_1 \sharp d_1 w_2 \sharp d_2 \ldots w_n \sharp$ where w_i is a word of $(a+b)^*$ and $d_i \in \{\sharp, \flat\}$, $\sigma(\rho)(c) = 2.\mathbb{P}_{\mathfrak{A}}(w_n)$ and $\sigma(\rho)(l) = 1 - 2.\mathbb{P}_{\mathfrak{A}}(w_n)$. For every other path, σ chooses the only available action: e. With this choice, for all $i \in \mathbb{N}$ the probability that d_i is equal to $\flat$ for a secret or a non secret path is equal to $\mathbb{P}_{\mathfrak{A}}(w_n)$. Therefore, for any finite path ρ, ${\mathsf{P}^{\mathsf{rel}}}_{\mathsf{M}(\mu_0)}(\mathsf{O}(\rho)) = 1/2$. Thus $Disc_{min}^{\mathsf{AA}}(\mathsf{M}(\mu_0)) = 0$.

Consequently, the minimal almost sure disclosure decision problem is undecidable. ◀